

PATTERNS OF GROUNDWATER AND SOIL MOISTURE VARIABILITY:  
HARD DATA, SOFT DATA AND DOMINANT CONTROLS

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

**Michael Rinderer**

aus Österreich

Promotionskomitee

**Prof. Dr. Jan Seibert (Vorsitz)**

**Dr. Ilja van Meerveld**

**Dr. Manfred Stähli**

**Dr. Stefan Gruber**

Zürich, 2015

#### The Author:

Michael Rinderer was born in 1977 in Bludenz, a town in the mountainous, western part of Austria. He studied Geography at the University of Innsbruck (A) and Aberdeen (UK) and wrote his diploma thesis on bedload transport and debris flow modeling in alpine torrents. After graduation, he worked for seven years as a consultant in engineering offices and a private research center in Innsbruck focusing on natural hazard management. He decided to come back to university and conducted his PhD research in the Hydrology & Climate Group of the Geography Department at the University of Zurich.

*To*

*Stephan Jenewein*

*my close friend, study mate and working colleague in  
Innsbruck who died in an avalanche on the 22<sup>nd</sup> of  
March 2009.*

*Stephan Jenewein, Stephan Senfter and I were mapping  
the Erlenbach catchment in September 2005 when the  
three of us were working together in Innsbruck on a  
project for assessing runoff and bedload transport in  
alpine torrents. I wish we could have continued our  
discussion on runoff coefficients and dominant runoff  
processes in the Erlenbach, started during the  
fieldwork, in the light of the outcome of this thesis.*





# Patterns of Groundwater and Soil Moisture Variability: Hard Data, Soft Data and Dominant Controls

Michael Rinderer

Doctoral Dissertation from the Department of Geography, University of Zurich

---

## Abstract

Soil moisture and groundwater storage are important for understanding and predicting rainfall-runoff processes in watersheds. Research over the last 100 years has revealed detailed process understanding of infiltration and subsurface flow processes, but most studies have been restricted to small plots or hillslopes. However, catchment-scale hydrological functioning is not necessarily dominated by the same factors that control the response on the plot- and hillslope scale. Spatial patterns, such as groundwater and soil moisture patterns in the case of this thesis, reflect the spatial organization of natural hydrological systems. These patterns can be analyzed in terms of connectivity between runoff generation areas in different parts of the catchments and the stream network to investigate functional relations between pattern connectivity and runoff response. Deciphering the dominant factors and small scale processes that control these patterns, helps to better understand how surface and/or subsurface flowpaths are established that efficiently contribute to runoff and dominate the runoff response at the catchment-scale. As time, effort and expenses for obtaining direct measurements of catchment-scale spatial patterns is high, qualitative or so called “soft” data can be a useful complement to quantitative or “hard” data.

In this thesis a new qualitative method called the “Boots & Trousers” method for mapping spatial patterns of soil moisture in humid environments and an adapted version for semi-arid conditions is proposed and systematically tested. Both methods are based on qualitative wetness indicators that one can see, feel or hear on the soil surface and are intuitive to local people from their every-day experience in outdoor activities (Switzerland) or crop growing and brick making (Tanzania). Both schemes were systematically tested to determine the correlation between qualitative wetness classes and quantitative differences in soil water content and for the agreement among classifications by different raters. It could be shown that the qualitative wetness classes reflect actual differences in volumetric water content. Neither experience, nor a certain level of education were a prerequisite for robust wetness classifications but a detailed introduction and training resulted in higher agreement among individual raters. The classifications for wet sampling points showed the highest agreement with the hard soil moisture data, while intermediate wetness classes seemed to be more difficult to assign. Some raters had a tendency to systematically rate specific wetness classes as too wet or too dry but when the raters were familiar with the application of the scheme, the mean offset was small and typically within the range of one wetness class.

In addition, the dominant topographic controls of median groundwater levels and groundwater response timing were investigated in a 20 ha pre-alpine catchment in the Alptal, Switzerland, with low permeability soils in order to predict spatial patterns of groundwater response for non-monitored sites. From the analysis of 51 groundwater monitoring sites and 133 rainfall events between 2010 and 2012 in the study catchment, it was shown that median groundwater levels were correlated to topographic indices including slope, curvature, Topographic Wetness Index (TWI) and upslope contributing area. The strength of correlation between groundwater levels and TWI decreased at the beginning of rainfall events, indicating large spatial differences in groundwater responses, and increased after peak flow, when groundwater levels could be considered as being spatially close to a steady state. Median groundwater response times were also correlated to topographic indices and decreased with increasing TWI for sites with  $TWI < 6$ , while wetter sites responded almost immediately to rainfall. Rainfall intensity was more important than antecedent moisture conditions for the slope of this functional relation.

The results of this thesis show that qualitative methods like the proposed “Boots & Trousers” method are reliable supplements of quantitative methods to capture the spatial variability in shallow soil moisture in different environments. They are fast to apply, require no experience, no measuring device and still provide robust and reliable results. They are therefore particularly suitable for mapping spatial soil moisture patterns in developing countries or remote areas.

The findings on the variability in groundwater highlight the importance of topography on median groundwater levels and groundwater response timing in mountain catchments with a low permeability soil suggesting differences in the dominant controls and runoff processes, compared to flatter watersheds with transmissive soils. The findings of this thesis are expected to be transferable to other catchments with similar character and therefore further our understanding to make predictions of soil moisture and groundwater storage and the runoff response of ungauged catchments.

Keywords: groundwater, soil moisture, spatial-temporal patterns, dominant controls, dominant runoff processes, mapping approach, inter-rater reliability, response timing, topographic controls, Topographic Wetness Index (TWI), subalpine catchment, rainfall threshold, antecedent wetness, soft data, hard data, Switzerland, Tanzania.

---

---

## Zusammenfassung

Bodenfeuchte und Grundwasserspeicher haben eine wichtige Bedeutung für das Prozessverständnis des Niederschlags-Abfluss Verhaltens in Einzugsgebieten. Trotz der Erkenntnisse zur Abflussbildung aus bisherigen Untersuchungen vorwiegend auf kleinen Versuchsflächen oder instrumentierten Hängen musste man erkennen, dass das hydrologische Verhalten ganzer Einzugsgebiete nicht notwendigerweise mit denselben dominanten Steuerfaktoren und Prozessen erklärt werden kann. Räumliche Muster, wie Muster der Bodenfeuchte und der Grundwasserstände in dieser Arbeit, können als funktionale Indikatoren dienen, welche die räumliche Struktur, in den an sich sehr heterogenen, natürlichen Einzugsgebieten, beschreiben. Die Identifikation dominanter Steuerfaktoren und Prozesse, welche diese räumlichen Muster generieren, ist ein vielversprechender Weg, um die Variabilität von Bodenfeuchte, Grundwasserständen, Bodenwasserflüssen und der damit verursachten Abflussreaktion auf der Skala eines ganzen Einzugsgebietes besser zu verstehen. Da Kosten, Zeit und Aufwand für das Erfassen räumlicher Muster innerhalb eines ganzen Einzugsgebietes sehr hoch sind, sind qualitative Daten, sogenannte „soft“ Daten eine vielversprechende Ergänzung zu herkömmlichen, quantitativen oder „harten“ Daten.

In diese Doktorarbeit wird eine neue, qualitative Methode, die sogenannte „Boots & Trousers“ Methode, für die Kartierung räumlicher Muster der Bodenfeuchte in feuchten Gebieten und eine leicht angepasste Variante für semi-aride Bedingungen präsentiert und getestet. Die „Boots & Trousers“ Methode basiert auf qualitativen Indikatoren, die man sehen, fühlen oder hören kann und deren Erkennung für die meisten Leute durch die alltägliche Erfahrung bei Freizeitaktivitäten intuitiv möglich ist. Die modifizierte Variante für semi-aride Gebiete bezieht sowohl die Erfahrung der lokalen Bevölkerung in Tansania bei der landwirtschaftlichen Feldarbeit, als auch die Erfahrung bei der Herstellung von Tonziegeln, mit ein. Beide Methoden wurden systematisch dahingehend untersucht, ob die qualitativen Feuchteklassen auch tatsächlichen Unterschieden im volumetrischen Bodenwassergehalt entsprechen und ob die Feuchteklassierungen unterschiedlicher Personen übereinstimmen oder stark divergieren. Es konnte gezeigt werden, dass die qualitativen Feuchteklassen Unterschiede im volumetrischen Bodenwassergehalt widerspiegeln. Weder Erfahrung noch Bildungsgrad waren Voraussetzung für robuste und zuverlässige Feuchteklassierungen, jedoch konnte durch eine gute Einführung oder gar ein Training der Handhabung der Methode ein besseres Klassierungsergebnis erzielt werden. Die Klassierungen von feuchten Standorten zeigten die beste Übereinstimmung, während es schwieriger schien, Standorte mit mittlerer Feuchte zuzuordnen. Einige Personen zeigten ein systematisches Verhalten, feuchte Standorte zu trocken und trockene Standorte zu feucht zu klassieren. Nachdem aber die Testpersonen die Methode und deren Handhabung verstanden hatten, war diese mittlere Abweichung in der Regel klein und zumeist geringer als eine Feuchteklasse.

In der Doktorarbeit wurden darüber hinaus die dominanten Steuerfaktoren der Grundwasserstände und deren Reaktionszeit in einem 20 ha grossen, voralpinen Einzugsgebiet im Alptal, Schweiz, untersucht, welches durch wenig durchlässige Böden charakterisiert ist. Die Analyse von 51 Grundwassermessstellen und 133 Niederschlagsereignissen im Zeitraum von 2010 bis 2012 in diesem Untersuchungsgebiet zeigte, dass die Grundwasserstände mit topographischen Indices wie der Hangneigung, der Krümmung, dem Topographischen Feuchte Index (TWI) und der Grösse des Einzugsgebietes oberhalb der Grundwassermessstelle korrelieren. Der Rang-Korrelationskoeffizient zwischen Grundwasserständen und TWI verringerte sich am Beginn eines Niederschlagsereignisses vermutlich aufgrund grosser, räumlicher Unterschiede in der Grundwasserreaktion, nahm aber nach der Ereignisspitze wieder deutlich zu, wenn sich alle Grundwasserstände in etwa ähnlich verhielten (Gleichgewichtszustand). Die Grundwasserreaktionszeiten waren ebenfalls mit den topographischen Indices korreliert. Für Standorte mit einem  $TWI < 6$  nahm die Grundwasserreaktionszeit mit steigendem TWI ab, während an den übrigen Standorten die Grundwasserstände beinahe zeitgleich mit dem Niederschlagsbeginn reagierten. Die Niederschlagsintensität beeinflusste die Steigung der Beziehung zwischen den Grundwasserständen und dem TWI stärker als die Vorfeuchtebedingungen.

Die Ergebnisse dieser Doktorarbeit zeigen, dass qualitative Methoden, wie die „Boots & Trousers“ Methode, eine zuverlässige Ergänzung zu quantitativen Methoden sein können, um räumliche Muster der Bodenfeuchte zu erfassen. Solche Methoden sind sehr schnell in der Anwendung, setzen keine Erfahrung voraus, benötigen kein Messgerät und können dennoch robuste und zuverlässige Ergebnisse liefern. Als solche sind qualitative Methoden unter anderem besonders für das Kartieren von Bodenfeuchtemustern in Entwicklungsländern und in entlegenen Gebieten geeignet.

Die Ergebnisse bezüglich der Grundwasservariabilität zeigen die Wichtigkeit der Topographie hinsichtlich der Grundwasserstände und der Grundwasserreaktionszeit in Gebirgseinzugsgebieten mit wenig durchlässigen Böden auf. Die Erkenntnisse lassen darauf schliessen, dass die dominanten Einflussfaktoren und Prozesse dort anders sind, als in flacheren Gebieten mit besser drainierten Böden, was für das Verständnis der Abflussreaktion und deren Vorhersage von zentraler Bedeutung ist.

Schlagworte: Grundwasser, Bodenfeuchte, räumlich-zeitliche Muster, dominante Einflussfaktoren, dominante Abflussprozesse, Kartiermethode, Inter-Rater Reliabilität, Reaktionszeit, topographische Einflussfaktoren, Topographischer Feuchte Index (TWI), subalpines Einzugsgebiet, Niederschlags-Schwellenwert, Vorfeuchtebedingungen, soft Daten, Schweiz, Tansania.

---

## List of Papers

- I. Rinderer M., Seibert J. (2012): Soil Information in Hydrologic Models: Hard Data, Soft Data, and the Dialog between Experimentalists and Modelers. In: *Hydropedology - Synergetic Interaction of Soil Science and Hydrology*, ed. H. Lin, p. 515–536, Elsevier B.V., Waltham.
- II. Rinderer M., Kollegger A., Fischer B. M. C., Stähli M., Seibert J. (2012): Sensing with boots and trousers - qualitative field observations of shallow soil moisture patterns, *Hydrological Processes*, 26, p. 4112–4120, doi:10.1002/hyp.9531.
- III. Rinderer M., Komakech H., Müller D., Seibert J. (in review): Qualitative soil moisture assessment in semi-arid Africa - The role of experience and training on inter-rater reliability, *Hydrology and Earth System Sciences*.
- IV. Rinderer M., van Meerveld H. J., Seibert J. (2014): Topographic controls on shallow groundwater levels in a steep, prealpine catchment: When are the TWI assumptions valid?, *Water Resources Research*, 50(7), p. 6067–6080, doi:10.1002/2013WR015009.
- V. Rinderer M., van Meerveld H. J., Stähli M., and Seibert J. (in review): Groundwater response timing in a prealpine catchment, *Hydrological Processes*.

## Author Contributions

I was responsible for developing the measuring concept in WS07 in the Alptal, getting legal permission from the local authority and landowner, installing the monitoring network, mapping the site characteristics and organizing and carrying out the repair and regular maintenance (2010, 2011, 2012). I received input on the measurement concept from Russell Smith, support on getting legal permission by Manfred Stähli (WSL) and field support during the installations in 2010 (for a complete list of helpers, see the acknowledgments). Mirjam Zehnder conducted slug & bail tests in WS07 and provided saturated hydraulic conductivity data for paper IV. I cleaned and processed all data and checked its plausibility, implemented and performed all statistical and GIS-analysis, designed all figures and had the lead responsibility for writing the manuscripts. The co-authors of the individual papers contributed to the writing by discussing, commenting and modifying my text.

Jan Seibert had the initial idea to use qualitative indicators for soil moisture assessment that Andrea Kollegger followed in her Msc-thesis supervised by Benjamin Fischer and myself. Andrea Kollegger provided dataset 3 for paper II. I developed the idea to systematically test the qualitative soil wetness classification scheme in terms of inter-rater reliability, correspondence of qualitative and quantitative soil wetness and designed and performed the test with students, farmers and experts in Switzerland and Tanzania. I received input from Hans Komakech and other Tanzanian partners on the qualitative soil classification scheme of semi-arid environments, its translation to Swahili and support in the organization of the two tests in Tanzania. Daniela Müller contributed to the TDR-, gravimetric sampling and soil analysis in Tanzania (see a complete list of other helpers in the acknowledgments).

<b>1. Introduction.....</b>	<b>11</b>
1.1 Spatial Patterns and their Dominant Controls .....	11
1.2 Hard Data – Soft Data .....	12
1.3 Qualitative Assessment of Soil Moisture Patterns.....	12
1.4 Dominant Controls of Groundwater Variability .....	13
<b>2. Thesis Objectives.....</b>	<b>13</b>
2.1 Dominant Controls and Dominant Runoff Processes .....	13
2.2 Qualitative Soil Moisture Assessment.....	14
2.3 Groundwater Variability.....	14
<b>3. Methods.....</b>	<b>14</b>
3.1 Qualitative Soil Moisture Assessment.....	14
3.1.1 Soil Wetness Classification Schemes .....	14
3.1.2 Study Sites, Datasets and Test Layout .....	15
3.1.3 Analytical Methods .....	15
3.2 Groundwater Variability.....	16
3.2.1 Study Catchment .....	16
3.2.2 Analytical Methods .....	16
<b>4. Results .....</b>	<b>17</b>
4.1 Qualitative Soil Moisture Assessment.....	17
4.1.1 Inter-Rater Reliability of Qualitative Soil Wetness Classifications .....	17
4.1.2 The Effect of Experience and Training on the Inter-Rater Reliability.....	18
4.2 Groundwater Variability.....	18
4.2.1 Median Groundwater Levels.....	18
4.2.2 Groundwater Response Timing.....	20
<b>5. Discussion.....</b>	<b>20</b>
5.1 Qualitative Soil Moisture Assessment.....	20
5.1.1 Reliability of the Boots & Trousers Method .....	20

5.1.2 <i>The Effect of Experience and Training on the Inter-Rater Reliability</i> .....	22
5.2 Groundwater Variability .....	22
5.2.1 <i>Median Groundwater Levels and Topographic Controls</i> .....	22
5.2.2 <i>Groundwater Response Timing and Topographic Controls</i> .....	23
<b>6. Conclusions</b> .....	<b>24</b>
6.1 Qualitative Soil Moisture Assessment .....	24
6.2 Groundwater Variability and Dominant Controls .....	25
<b>7. Outlook</b> .....	<b>25</b>
<b>8. Acknowledgments</b> .....	<b>26</b>
<b>9. References</b> .....	<b>28</b>

## 1. Introduction

### 1.1 Spatial Patterns and their Dominant Controls

For understanding and predicting rainfall-runoff processes in watersheds, the soil water storage plays an important role. The state of the unsaturated zone storage – the soil moisture content – is important for interactions with plants in terms of moisture uptake, interactions with the atmosphere either in terms of moisture losses due to evaporation or in terms of infiltration of rainfall and subsequent percolation to the saturated zone. The saturated zone storage – the groundwater – and the saturated hydraulic conductivity determine the efficiency, with which groundwater can be redistributed by subsurface flow, establish connections to the channel network and therefore contribute to streamflow. The saturated and unsaturated soil storage also determine the amount and duration of water that is stored in a catchment.

The streamflow response to rainfall or snowmelt in small headwater catchments is often non-linear and threshold-like and a lot of research has been undertaken to better understand the underlying mechanism (Hewlett & Hibbert 1967; Whipkey 1965; Dunne & Black 1970; Mosley 1979; Sklash & Farvolden 1979; Sklash et al. 1986; Sidle et al. 2000; McDonnell 1990). Traditionally these studies have followed a reductionist approach, trying to study and model hydrological processes in more and more detail, which often only allowed for processes at selected points, small plots or hillslopes to be investigated. These studies have revealed interesting details about the complexity of natural systems but not necessarily helped to

better understand and predict runoff response at the scale of an entire catchment as heterogeneity in catchment properties causes process variability at different scales (McDonnell et al. 2007; Tetzlaff et al. 2008).

Dooge (1986) was among the first to express the need for macro-scale laws in hydrology and suggested the search for scale invariance and scale dependence of landscape properties and watershed response. He stated that catchments are “... complex systems with some degree of organization” (Dooge, 1986:48). A possible way forward is, therefore, to analyze landscape heterogeneity and process complexity in terms of hydrological function that dominates the behavior at the catchment scale and to identify the underlying organizing principles (Sivapalan 2005).

Spatial patterns, such as soil moisture and groundwater patterns, reflect the spatial organization of hydrological processes. They are functional characteristics or traits – an ecological concept (Darwin 1859) also transferable to hydrology – that impact the system response at a higher organizational level or scale by enabling/disabling collection, storage and release of water and the associated flows. Relations between structural traits and system functioning allow the response of a system to be predicted (Violle et al. 2007; Chaturvedi et al. 2011). Spatial patterns can be investigated in terms of hydrologic connectivity of active runoff generation areas and the stream network and can be related to the runoff response at the catchment scale. Hydrologic connectivity is defined as the degree to which surface and/or subsurface flow of different parts of a catchment are linked to the stream network (Ali & Roy 2009). We can gain better understanding of the non-linear functioning of a catchment, if we understand the small-scale processes and dominant controls that cause the spatio-

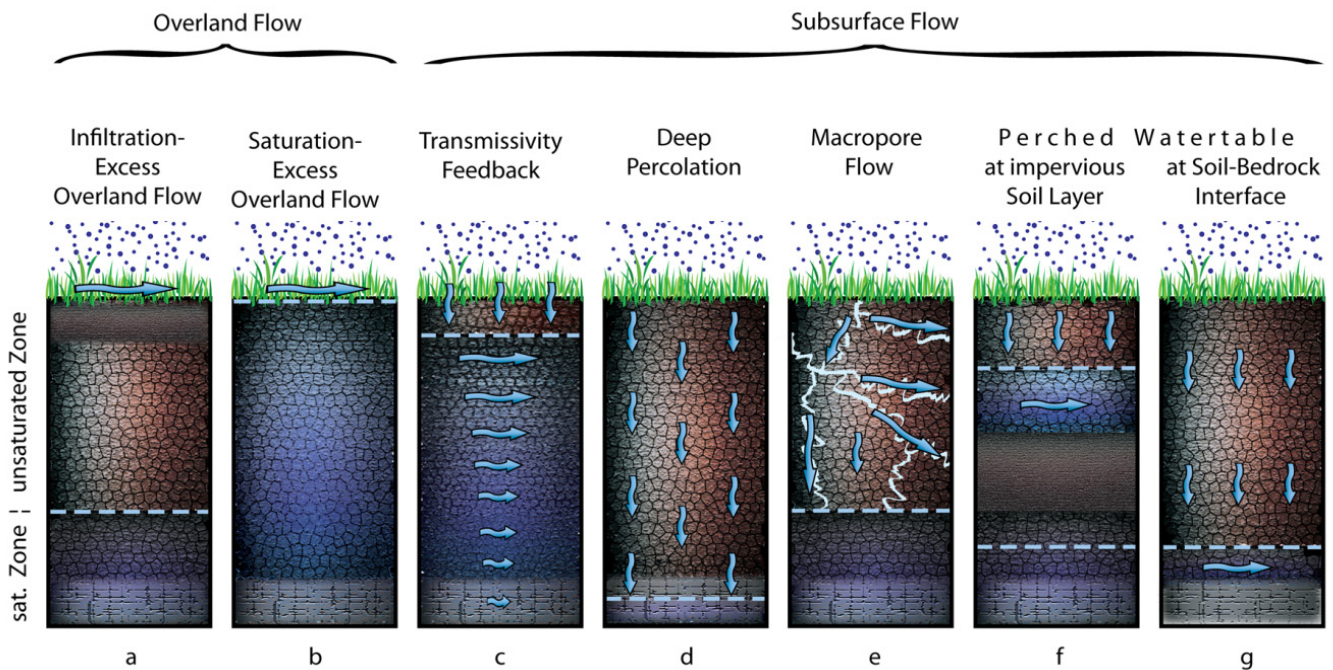


Figure 1: Different types of dominant runoff processes that can connect runoff source areas with the stream network, resulting in runoff response at the catchment outlet (figure taken from paper I).

temporal patterns of groundwater response. These groundwater response patterns link up in a self organizing manner to efficient continuous pathways that dominate the runoff response at a larger scale (McDonnell et al. 2007).

There are numerous pathways in the soil profile and at the soil surface that can potentially connect active runoff generation areas and the stream network (Figure 1). Instead of describing them all in full detail, it is more informative to identify the dominant flow pathways and associated runoff processes (Grayson & Blöschl 2000). Two types of decision schemes with objective criteria and defined procedures can be used to infer dominant runoff processes from site characteristic. The bottom-up approaches (e.g., Scherrer & Naef 2003; Schmocker-Fackel et al. 2007) are based on detailed soil survey information and artificial sprinkling experiments at the point or plot-scale to identify dominant runoff processes for representative sites in a catchment. These dominant runoff processes are then assigned to other areas in the catchment with similar properties based on the distribution of soil type, land-use and vegetation in a catchment (e.g., Peschke et al. 1999; Schüler 2005; Waldenmeyer 2003). Top-down approaches first identify hydrologically homogeneous landscape units based on coarse maps (e.g., soil maps, geological maps and land-use maps) and remote sensing data and then assign runoff processes that are expected to dominate their hydrological response to these units (Boorman et al. 1995; Uhlenbrook 2003; Tilch et al. 2006).

How these dominant runoff processes establish and maintain a hydraulic connection to the stream network is expected to be an emergent property, typical for each individual catchment but governed by the same small-scale processes. Identification of the dominant controls and dominant runoff processes that result in the spatial groundwater and soil moisture patterns allows groundwater and soil moisture storage to be predicted at sites or catchments with no measurements, and therefore enables a better understanding of catchment functioning across similar catchments.

### *1.2 Hard Data – Soft Data*

The most direct way to identify dominant runoff processes and their controls is by measuring subsurface flow and soil hydraulic properties during field experiments and monitoring campaigns (Scherrer & Naef, 2003, paper IV & V). The effort and cost for such investigations are high but these experiments result in quantitative or so called “hard” data, which can be analyzed directly with quantitative methods and models. Information about soils and dominant runoff processes can also be obtained from “soft” data, which is qualitative, semi-quantitative or categorical information (Seibert & McDonnell 2002; paper I). This information needs the experts’ interpretation or a critical quality assessment before it can be analyzed with quantitative methods. Typical examples are historic flood records (Schmocker-Fackel & Naef 2010) or mean transit times that are derived from tracer experiments (Soulsby & Tetzlaff 2008). Soft data can be indirect measurements of

the value of interest (e.g., thermal differences in an infra red image reflect differences in soil moisture; Pfister et al. 2010). Soft data also include personal experience also called tacit knowledge that someone gained in his/her daily life as a farmer, forester or hiker (paper II & III).

A lot of information about dominant controls and dominant runoff processes already exists but is not fully exploited in data analysis and hydrological modeling so far (paper I). Soil classes and their taxonomy give indirect information about the degree, duration and nature of soil saturation (e.g., Gleysol versus Pseudogley; see also the HOST classification by Boorman et al. 1995 and (Gillin et al. 2014). Similar information can be derived from natural drainage classes, water-state annual patterns and inundation classes that are documented in soil survey databases (Soil Survey Division Staff 1993). If a soil profile has been documented, the sequence of the different soil horizons and their soil hydraulic properties as well as hydromorphic features (Fe and Mn precipitation), can provide information on the dynamics of the groundwater table and associated subsurface runoff processes (Soil Survey Staff 2010; Lin et al. 2008). Other variables like topography allow dominant soil moisture states and associated dominant runoff processes to be inferred, as they influence the drainage of water. This shows that there is a wealth of information already included in existing soil survey databases that is not yet fully exploited in hydrological terms due to its qualitative nature.

### *1.3 Qualitative Assessment of Soil Moisture Patterns*

In humid-temperate climate conditions, spatial patterns of soil moisture have been used to identify local and non-local controls on the redistribution of water in the unsaturated soil profile (Grayson et al. 1997; Blume et al. 2009; Western & Grayson 2000). Maps of soil moisture, including saturated areas, are potentially valuable for inferring hydrological connectivity (Western et al. 2001; McNamara et al. 2005; Ali et al. 2010) and for validating and calibrating models (Beven & Kirkby 1979; Blazkova et al. 2002; Güntner et al. 2004).

In semi-arid conditions soil moisture is of key-importance for crop survival and for allocating irrigation water based on the differences in the soil moisture state of the fields belonging to individual farmers in a community. Common techniques for measuring soil moisture are often time consuming and/or rely on expensive equipment that is usually not available to farming communities in developing countries. Therefore decision making on allocation of water resources during periods of water scarcity currently depends on the decision of community leaders, whose assessment might be disputed. A more systematic way of soil wetness assessment based on defined criteria would relieve pressure on community leaders and assure transparency in decision making and therefore avoid conflicts among farmers.

Qualitative methods have been shown to be useful complements to quantitative measurement techniques in a number of field applications in soil science, risk assessment



and ecology. They are based on qualitative indicators that one can identify through sight, sound or touch and that are related to quantitative properties of interest, like the texture of a soil sample (Thien 1979), the strength of a snow pack (De Quervain 1950; cited in Pielmeier & Schneebeil 2003) or the water quality of a stream (Metcalf-Smith 1994).

In hydrology, qualitative indicators have been used for mapping saturated areas (Dunne & Black 1970; Dunne et al. 1975). Soil hydromorphic features that are visible from a soil profile or vegetation patterns that can be mapped in the field are useful indicators of prevailing soil moisture conditions as well (Ellenberg et al. 1991; Quinn et al. 1998; Kulasova et al. 2014; Gillin et al. 2014; paper I). However these methods do not allow different grades of soil wetness or changes in soil wetness over time to be captured. The “spade diagnosis” method for an applied soil texture examination in the field, includes a soil wetness classification scheme with five qualitative wetness classes (Görbing & Sekera 1947). The Natural Resources Conservation Service of the United States Department of Agriculture (1998) published guidelines for estimating soil moisture by feel and appearance for four different soil types and different soil moisture content. Blazkova et al. (2002) defined a classification scheme based on five qualitative wetness classes but they were mainly interested in the three wettest ones to assess soil saturation. It remains unclear if these methods were tested and how reliably they can be applied by different people.

#### *1.4 Dominant Controls of Groundwater Variability*

In steep mountain headwater catchments, shallow groundwater can respond quickly to rainfall because alpine soils are typically thin and gradients are steep. In order to predict median groundwater levels and groundwater response timing, a better understanding of the factors that control the spatial variability in groundwater levels and groundwater response is important (McGlynn & McDonnell 2003; McDonnell et al. 2007).

Previous studies on groundwater response have revealed that the percentage of groundwater wells that showed a response during individual rainfall events depended on a rainfall threshold and was correlated to total event precipitation and storm duration, but not to rainfall intensity and antecedent conditions (Penna et al. 2014; Dhakal & Sullivan 2014; Fannin et al. 2000). Furthermore, studies showed that the groundwater response timing was related to landform (Detty & McGuire 2010), distance to the stream channel network (Seibert et al. 2003; Rodhe & Seibert 2011; Haught & van Meerveld 2011) and thickness of the soil or the topography of the bedrock (Penna et al. 2014; Tromp-van Meerveld & McDonnell 2006). But the observations reported in the literature are ambiguous in terms of spatial groundwater response behavior. In some catchments, groundwater levels close to the stream responded first, while upslope sites responses were delayed (Haught & van Meerveld 2011). In contrast, other studies reported the shortest groundwater response times in the upper parts of the hillslopes (Penna et al. 2014) or could not

find a systematic pattern in the timing of the groundwater response (Lana-Renault et al. 2014).

These partly contradicting observations have made it difficult to generalize groundwater response behavior. Under wet environmental conditions (Anderson & Burt 1978; Burt & Butcher 1985; Lana-Renault et al. 2014), steep terrain (Penna et al. 2014) or shallow groundwater tables (Troch et al. 1993), variability in groundwater responses were related to topography. Under dry conditions (Detty & McGuire 2010), flat terrain (Barling et al. 1994) and especially in permeable soils (Seibert et al. 1997; Dhakal & Sullivan 2014; Anderson et al. 2010), the relation between the groundwater response and topography is not yet clear.

## **2. Thesis Objectives**

This thesis contributes to a better understanding of the dominant controls on the spatial variation in groundwater and soil moisture using hard and soft data. New qualitative methods for mapping the spatial patterns of soil moisture are proposed and tested and the dominant controls of median groundwater levels and groundwater response timing are investigated. This allows predictions of spatial patterns of storage and hydrological connectivity to be made, to better understand the runoff response at the catchment scale.

### *2.1 Dominant Controls and Dominant Runoff Processes*

Paper I describes the role of soil in terms of modulating rainfall inputs to runoff. Experimentalists gain field experience – called tacit knowledge – by observing and documenting runoff processes and hydric soil indicators. Modelers on the other hand try to break down natural process complexity into simplified process descriptions. The challenge for both is to identify the dominant controls of hydrologic functioning to better understand the often non-linear runoff response at the catchment scale. Paper I therefore answers the following questions:

- 1) How are subsurface runoff processes represented in models of different complexity, ranging from simple conceptual models to more complex physically-based models?
- 2) How can catchment-scale models be parameterized using point-scale measurements and existing model approaches originally developed for small scales (e.g. a soil column)?
- 3) Which information can be gained from soil survey methods, including mapping approaches of hydric soil indicators?
- 4) Can decision schemes be useful to indicate dominant runoff processes in an objective way?

The soft data concept is seen as a possible way forward to enhance the dialog between the experimentalists and modelers. Soft data can be made useful for modeling by applying fuzzy-logic based functions to evaluate the degree

to which model simulations compare to the experimentalists' field experience.

## 2.2 Qualitative Soil Moisture Assessment

In paper II, the “Boots & Trousers” method, a new qualitative wetness classification scheme for capturing shallow soil moisture differences in humid environmental conditions is presented as a novel type of soft data. It is based on qualitative wetness indicators that one can see, feel or hear. A modified version for semi-arid environmental conditions is proposed in paper III, incorporating the local peoples' experience in Tanzania in terms of the soil wetness that is optimal for seeding crops and brick making. Both methods are seen as a supplement to quantitative methods. The schemes are systematically tested for the agreement between the qualitative wetness classes and the quantitative differences in soil water content and for the agreement among classifications by different raters.

In particular, the following questions are addressed:

- 1) Is a qualitative wetness classification scheme, such as the “Boots & Trousers” method or its adaptation to semi-arid conditions, capable of reliably capturing differences in shallow soil moisture conditions?
- 2) Do the qualitative wetness classes reflect actual differences in volumetric water content of the study sites?
- 3) To what extent is the qualitative wetness classification scheme subjective?
- 4) Are there differences in the agreement among classification of the sampling points with different wetness?
- 5) Do some individual raters systematically classify the sampling points as too wet or too dry compared to the rest of the group?
- 6) Does the agreement of qualitative wetness classifications depend on the level of education or experience of the rater?
- 7) Does the way in which the classification scheme is introduced to the participants and how they are trained affect the variability in wetness class assignments by the raters?

## 2.3 Groundwater Variability

Despite the knowledge gained on groundwater dynamics in previous hillslopes studies, we still know relatively little about catchment-scale groundwater dynamics and the dominant controls on groundwater responses in steep mountain environments with low permeability soils. One might expect the groundwater levels to be more responsive to rainfall in these environments because of the lower storage deficit, low drainable porosity and low hydraulic conductivity of the mineral soil. The topography is expected to be a dominant control on groundwater levels as the gravitational potential in mountain headwaters is high. These hypothesis are tested in a pre-alpine headwater

catchment in Switzerland by analyzing the spatial variability in the median groundwater levels and the groundwater response timing and correlating it to topographic indices and rainfall and antecedent wetness characteristics. Furthermore it was also tested if the groundwater level variations can be approximated by a series of steady-state successions in steep mountain headwater catchments with low permeability soils. The following specific research questions are addressed:

- 1) To what extent does topography control median groundwater levels and the groundwater response timing in a catchment with low permeability soils?
- 2) Are there differences in the correlation of median groundwater levels and groundwater response timing with local and upslope topographic characteristics?
- 3) Is there a rainfall threshold for groundwater response initiation and if so, does this threshold depend on topography?
- 4) How do antecedent soil wetness conditions and rainfall intensity influence the timing of the groundwater response?
- 5) Does the correlation between topography and groundwater levels vary over time?

## 3. Methods

### 3.1 Qualitative Soil Moisture Assessment

#### 3.1.1 Soil Wetness Classification Schemes

The wetness classification scheme presented in paper II is based on seven qualitatively defined classes: At a location classified as class 1, the trousers of a person would stay dry when sitting down on the ground. For a site classified as class 2, trousers would get moist after several minutes, at class 3, trousers would get wet after several minutes and at class 4, trousers would get wet immediately when sitting down on the ground. If one could hear a squelchy noise when stepping on the ground with a boot, the spot would be classified as class 5, if in addition water would squeeze out of the soil, it would be classified as class 6, and if one could see water ponding on the soil surface, it would be classified as class 7. Obviously, it is not intended that a rater actually sits down on the surface each time to assess the wetness conditions but it is rather assumed that most people have some experience and know, or can imagine, whether they would stay dry or get wet if they were to sit on the ground. As vegetation and litter layers potentially influence the rater's class assignments they should be bent aside or be removed.

The qualitative soil wetness classification scheme presented in paper III is a modification of the “Boots & Trousers method (paper II) for semi-arid environments. It is adapted to the local peoples' experience in terms of soil wetness that is optimal for seeding crops and brick making in Tanzania. It consists of seven qualitative soil wetness classes: The driest class 1 is called “very dry – dust dry” for which one cannot see or feel any moisture in the soil at the

soil surface. Class 2 is characterized by a soil sample which is dry but has a moist look. Class 3 is slightly drier than the optimal seeding conditions, class 4 indicates optimal wetness for plant seeding, class 5 is optimal for making bricks, and class 6 being too wet to form a brick. For the wettest class 7 one can see water ponding on the soil surface. It is not intended to tie optimal seeding conditions to a specific crop or actually form a brick to test its stability but rather to reflect farmers' experience in imagining these conditions from their every-day life.

### 3.1.2 Study Sites, Datasets and Test Layout

The Erlenbach catchment (70 ha) in the Alptal, a pre-alpine valley about 40 km southeast of Zurich (Switzerland) was chosen to test the Boots & Trousers method (paper II). In the catchment, 0.5 to 2 m deep umbric or mollic Gleysols with Muck Humus and Mor Humus topsoils have formed on top of a marly parent material (Schleppi et al. 1998). Due to humid environmental conditions, with 2300 mm precipitation a year (Feyen et al. 1999), moor landscapes can be found. Non-forested locations are generally classified as wetter sampling points. Drier sampling points were either located on steeper, convex slopes or in areas with a light forest cover (predominantly Norway Spruce).

In May 2011 a test with 20 master students (dataset 1) was organized during which they were asked to classify 52 sampling points of different wetness arranged in a random order and marked with a flag. Half of the students (group1) performed the test in the morning and half of them (group2) in the afternoon, when the soil had dried up. All participants were given a 5 minute basic introduction, but no training.

A different set of 45 sampling points (dataset 2) was selected to compare qualitative soil wetness classifications with the corresponding volumetric soil water content. Soil samples were taken with a 100 cm<sup>3</sup> steel cylinder in 10 cm depth and soil wetness was classified by three experts using the qualitative scheme. In addition, TDR measurements and qualitative wetness classifications of 100 marked sampling points in the Erlenbach catchment (dataset 3), were collected during eight sampling campaigns between August and October 2010 (Kollegger 2010). A quality check of the TDR data reduced dataset 3 to 454 data records of TDR- and qualitative classification pairs.

The wetness classification scheme for semi-arid conditions was tested in the two farming villages Mungushi and Kichangani, in the upper Pangani basin, ca. 25 km southeast of Arusha / Tanzania. Haplic Andosols (loamic, fluvic), with a texture classified as Clay Loam (sand: 35 % silt: 28%, clay: 37%) dominate the area. Soils are fertile and heavily used for growing crops, mainly beans and corn. Due to a limited amount of rainfall (less than 600 mm/year) (Komakech & Van der Zaag 2011) falling mainly during the rainy season, agriculture in this region depends on flood irrigation during the rest of the year.

To test the wetness classification scheme we performed two experiments. The first test in April 2014 was organized in the Mungushi village where 40 sampling points of

different wetness were marked with flags. The test involved 40 people, namely 14 farmers, 14 master students, 11 experts (PhD and professors). All participants were given a brief introduction, of about 5 minutes, to the wetness classification scheme either in Swahili or English and then were asked to individually classify the marked sites of different wetness along the marked parcours. Half of the farmers ( $F_{trained}$ ) and 1/3 of the students ( $S_{trained}$ ) received an additional training (~10 min) in which they were shown representative sites of wetness classes 1, 4, and 7. Farmers and students with a basic introduction are called  $F_{basic}$  and  $S_{basic}$ , and when referring to all of the farmers, students and experts we use the expressions  $F_{all}$ ,  $S_{all}$  and  $E_{all}$ . The assessment form used in April 2014 consisted of a matrix on an A4 paper (landscape format) with the number of the sampling sites appearing as rows and the wetness classes as columns. Participants were asked to tick the appropriate cell corresponding to their judgment of the soil moisture condition of a particular site.

In June 2014 a similar test with 18 farmers and 7 experts was organized in the neighboring village of Kichangani (42 sampling points) during which the participants were given a longer introduction (~20 min) and better training (~30 min). In addition a new layout of the assessment form with pre-labeled sites and one column for assigning the wetness class number was tested. The flags were better labeled to prevent potential misinterpretation between the number of the site and the number of wetness class to assign. The wetness classification scheme remained the same. During both tests in April and in June, volumetric water content was also measured using the gravimetric method.

### 3.1.3 Analytical Methods

To assess the agreement of qualitative wetness classifications among farmers, students and experts, the frequency distribution of classification differences relative to the median of classifications of all group members, was analyzed. This was done by including all sampling points (overall agreement), for sampling points individually (wetness class specific agreement) and for individual persons (systematic bias). In paper II the mode was chosen as a reference as it represents the wetness class that most raters had agreed on. In paper III reviewers argued that the median was a more robust reference particularly when two or more classes have similar assignment frequencies. In this case the assignment of a single rater can change the mode of all classifications while the median is most likely less affected. Both, the median and the mode are not affected by outliers.

Krippendorff's Alpha (Krippendorff 2004) and Cohen's Kappa (Cohen 1960) were calculated to statistically assess the inter-rater reliability among raters. Krippendorff's Alpha is a measure to assess the degree of agreement within a group of raters and Cohen's Kappa (CK) between two raters, or, in our case, each individual rater and a reference. If there is no agreement among rates other than what would be expected by chance, KA and CK equals zero and if the raters agree perfectly, KA and CK would theoretically equal one. However for CK the maximum

attainable CK value (CK<sub>max</sub>) is normally smaller than one, as the frequency of class assignments between two raters is normally not equal. So kappa values were interpreted as the ratio between CK/CK<sub>max</sub> (Sim & Wright 2005). In this thesis KA and CK/CK<sub>max</sub> are given as percentages.

### 3.2 Groundwater Variability

#### 3.2.1 Study Catchment

The 20 ha study catchment is located in the Alptal, a pre-alpine valley about 40 km southeast of Zurich, Switzerland (Figure 2). It is a neighboring catchment of the Erlenbach (see paper II) and is not affected by anthropogenic drainage ditches. It is referred to here as WS07, as in Fischer et al. (in review). The catchment extends from 1270 m asl. to 1650 m asl., is steep (average slope: 35%) and has a distinct small scale topography and a dense drainage network. Due to the high mean annual precipitation of 2300 mm/year (Feyen et al. 1999) and low permeable soils (Gleysols) and bedrock (Flysch) (Schleppi et al. 1998), moor landscapes have formed with wet grassland growing in flat or concave parts of the catchment and open coniferous forest stands on steeper slopes and ridge-sites.

Fifty-one groundwater monitoring sites in seven nested sub-catchments (C1 to C7) were installed based on a stratified sampling procedure to cover the range of topographic positions, soil types and vegetation in the experimental catchments. The monitoring sites included 8 ridge site, 22 midslope and 21 footslope or depression sites;

20 sites were forested and 31 were located in grassland. The well depth varied between 0.5 to 2 m. Water levels were measured in the wells between September 2010 and the end of November 2012. The discharge at the outlet of each sub-catchment was measured continuously and climatologic variables were recorded at a long-term weather station 1 km from the experimental catchment. For the analyses of the groundwater timing, we selected 133 rainfall events during the snow-free periods and further classified them into four rainfall event types according to mean rainfall intensity and 3 day antecedent rainfall.

#### 3.2.2 Analytical Methods

Several topographic indices based on a 6 by 6 meter Digital Terrain Model (DTM) were calculated that were expected to represent dominant controls of groundwater levels. *Local controls* are defined as properties that characterize the monitoring site itself and *upslope controls* as the properties that characterize the upslope contributing area. Site characteristics selected for paper IV and paper V were: local slope gradient (Tarboton 1997), local curvature (Evans 1980; Travis et al. 1975), TWI (Beven & Kirkby 1979), upslope contributing area, mean slope, mean curvature and mean TWI of the upslope contributing area. To quantify the relation between topographic characteristics and median event timing characteristics the Spearman rank correlation coefficient ( $r_s$ ) was used.

The Spearman rank correlation coefficient ( $r_s$ ) (Spearman 1904) was chosen in paper IV and paper V as a measure to quantify the relation between groundwater level

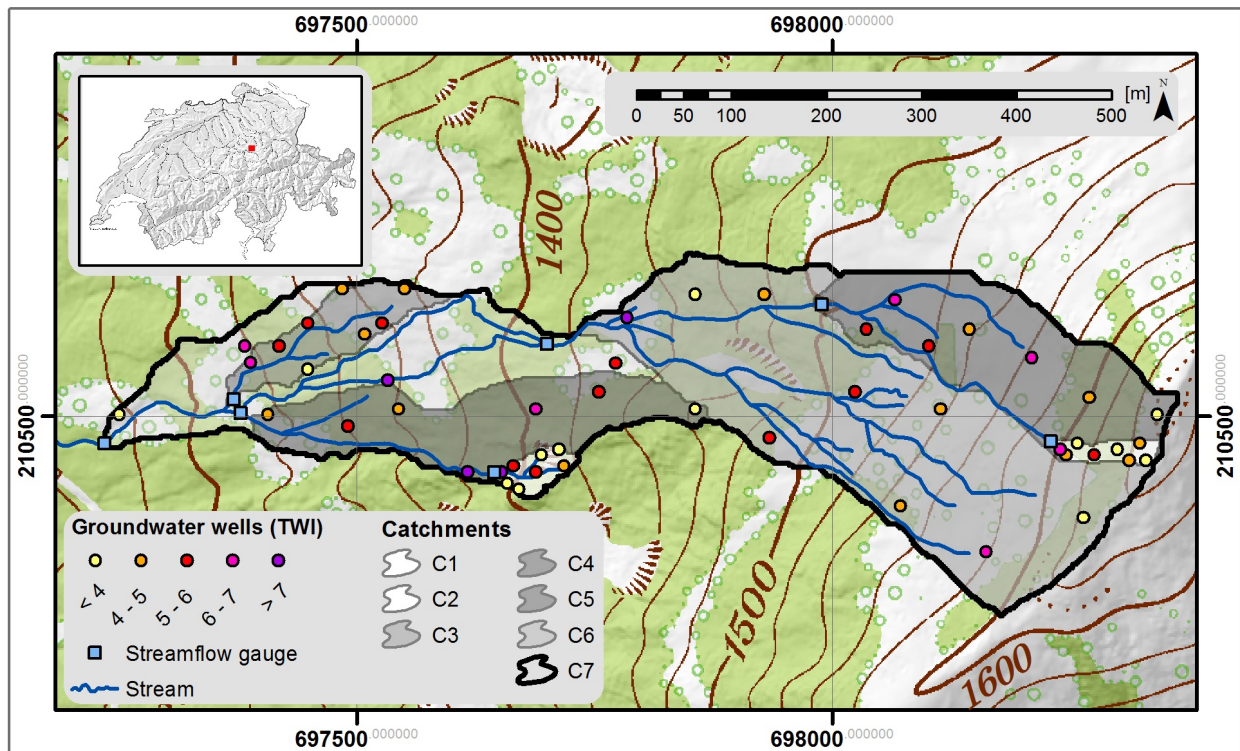


Figure 2: Map of the experimental catchment showing the seven nested sub-catchments with a streamflow gauging station at each outlet and the location of the 51 spatially distributed groundwater wells. Groundwater wells are color-coded according to the Topographic Wetness Index (TWI), (Background-topographic map: Swiss topo, 123456789; figure taken from paper V).

characteristics and the topographic indices. In paper IV the correlation analysis was based on median groundwater levels because they are less affected by censored data, e.g., when the groundwater level falls below the bottom of the groundwater well. Groundwater levels were scaled by the soil depth (1 = water level at the soil surface, 0 = dry well) to compensate for differences in soil depth and well depth among monitoring sites, respectively.

For the assessment of the change in correlation between groundwater levels and TWI,  $r_s$  was calculated based on hourly streamflow data and the relative change in streamflow ( $dQ/Q$ ) of sub-catchment C5. Streamflow was assumed to be an indicator of the system state and C5 was used because it provided the most complete runoff series. Data points were further classified into the *growing season* from the beginning of June until the end of September with frequent rainfall events, the *dormant season* between the beginning of October and the end of January and *spring*, including snowmelt between the beginning of February and the end of May.

For the timing analyses the *time to rise*, ( $t_{rise}$ ) was considered which is defined here as the time lag of groundwater rise relative to the start of the rainfall event. The *time to peak*, ( $t_{peakP}$ ) was calculated as the time lag between the centroid of each rainfall event and the 95% of total rise in groundwater level. The *groundwater peak duration* ( $t_{dur}$ ) was calculated as the time lag between the time that the water level had risen to 95% of the maximum water level on the rising limb of the groundwater hydrograph and the corresponding point of time on the falling limb (called 95% recession). The *duration of the recession* ( $t_{rec}$ ) was defined as the time between 95% of the rise and 20% of the rise on the falling limb of the groundwater hydrograph. In order to investigate synchronicity between groundwater and streamflow peaks  $t_{peakQ}$  was calculated as the time lag between the 95% peak groundwater level and 95% peak discharge at the catchment outlet. The rainfall threshold to initiate groundwater response ( $P_{rise}$ ) was analyzed, which is the sum of rainfall that fell between the start of the rainfall and  $t_{rise}$ .

## 4. Results

### 4.1 Qualitative Soil Moisture Assessment

#### 4.1.1 Inter-Rater Reliability of Qualitative Soil Wetness Classifications

In 72% and 67% of all soil wetness classifications members of group1 and group2 independently assigned the same wetness class and in about 96% and 92% of all classifications they agreed or were off by not more than one class. An over- or underestimation of wetness by more than two classes occurred in 3 cases (0.6%) and 5 cases (1%) among members of group1 and group2, respectively.

Raters of both groups agreed to a large extent in wetness class assignment for wet to intermediate sampling points. This is expressed by a narrow frequency distribution with

most or almost all assignments for a single sampling point falling within the same wetness class (for group2 see Figure 3a). A small spread of class assignments could be identified for class 5 to 7 (for both groups) and for class 4 of group1 while wetness classes 2 and 3 showed a larger spread of class assignments.

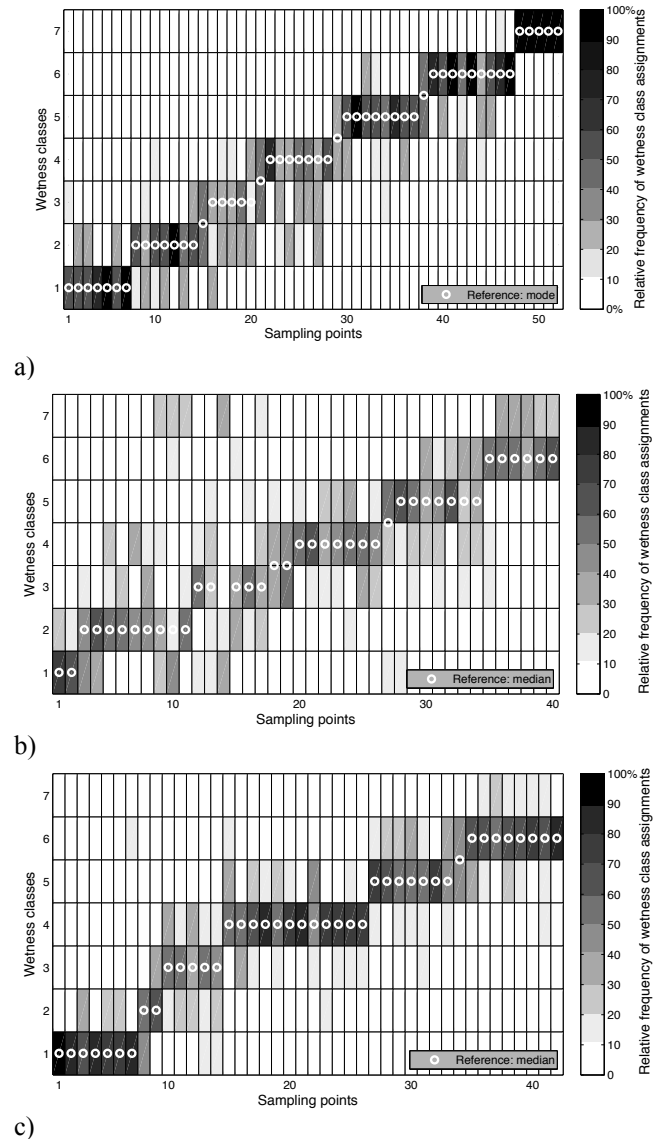


Figure 3: Spread of classification assignments for sampling points of individual soil wetness classes by a) students (group2) in Switzerland during the test in May 2011 (paper II), b) farmers in Tanzania during the test in April 2014 and c) farmers during the test in June 2014 with a better introduction (paper III). (grey-shades: relative frequency of wetness class assignments for each of the sampling points). The sampling points were distributed in random order of wetness classes in the field experiment but are ordered here for graphical clarity (figure taken from paper II and paper III).

The Kappa statistic corroborated these results as the median  $CK/CK_{max}$  of group1 and group2 was 81% and 73%, respectively. Krippendorff's Alpha was 84% for group1 and 87% for group2.

The majority of raters did not show a systematic tendency to classify all sites as too dry or too wet. For

individual wetness classes there was a tendency for individual raters to classify wet to intermediate sites as too dry and the driest sites as too wet but the mean difference was normally smaller than one wetness class. The frequency distribution for class 7 was the narrowest.

In general, the qualitative wetness classes reflected differences in the median volumetric water content determined from the gravimetric samples (dataset 2) and the TDR measurements (dataset 3). However for the intermediate wetness classes 2, 3, 4 and for the wettest classes 6 and 7 of dataset 2 the median volumetric water content was similar. This was also true for sampling points of class 6 and class 7 in dataset 3 but the median volumetric water content of the intermediate wetness classes in dataset 3 was more distinct compared to dataset 2. All classes showed a large inter quartile range (IQR), in particular, class 2 of dataset 2.

#### 4.1.2 The Effect of Experience and Training on the Inter-Rater Reliability

In terms of the role of experience in crop growing and level of education on the agreement of wetness classifications using the qualitative scheme for semi-arid conditions in Tanzania, the first test in April showed that the  $F_{all}$  had a lower degree of agreement than  $S_{all}$  and  $E_{all}$ . In about 46% of all cases classified by  $F_{all}$  they agreed and independently assigned the same wetness class while  $S_{all}$  and  $E_{all}$  agreed on the same wetness class in 60% and 59% respectively. 22 assignments by  $F_{all}$  were off by four or more classes while it was only 1 and 2 for  $S_{all}$  and  $E_{all}$ , respectively (for farmers see Figure 3b).

The difference in the degree of agreement between  $F_{all}$ ,  $S_{all}$  and  $E_{all}$  during the test in April was also evident from the inter-rater reliability statistics. The Krippendorff Alpha (KA) value for  $F_{all}$  (42%) was half of KA of  $S_{all}$  (83%) and  $E_{all}$  (82%) during the test in April. The median  $CK/CK_{max}$  also differed between  $F_{all}$ ,  $S_{all}$  and  $E_{all}$  (43%, 65 % and 67%); Figure 4).

During the second test in June the agreement of class assignments among  $F_{all}$  was higher and exceeded even the agreement among  $E_{all}$  : In about 66% of all cases  $F_{all}$  independently assigned the same wetness class and  $E_{all}$  in 59%. No expert was off the group median by more than two wetness classes during the second test but  $F_{all}$  were off by more than two classes in 13 cases (2%) of all classifications (for farmers see Figure 3c).

During the second test in June,  $F_{all}$  achieved a similar inter-rater reliability to  $E_{all}$  (no student raters during the test in June). KA of  $F_{all}$  (76%) was more similar to KA of  $E_{all}$  (84%) and the median of  $CK/CK_{max}$  of  $F_{all}$  (75%) even exceeded that of  $E_{all}$  (59%) during the second test in June (Figure 4). The IQR of  $CK/CK_{max}$  for  $F_{all}$  during the second test was almost half the IQR of the first test.

In terms of the role of training on how to apply the wetness classification scheme, it became evident that  $S_{trained}$  during the test in April and  $F_{trained}$  during the test in June had a higher interrater reliability (KA and

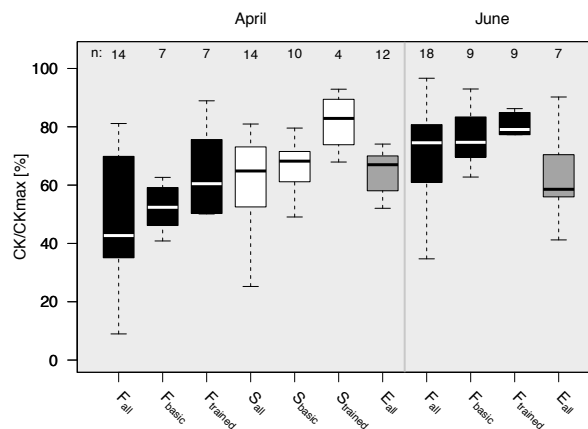


Figure 4: Inter-rater reliability among members of Tanzanian farmers (F), students (S) and experts (E) tested in April and June 2014 expressed as a ratio between  $CK/CK_{max}$ . (Subscripts "basic" indicates the sub-group with only basic introduction, "trained" indicates the sub-group with more detailed training, "all" indicates that both subgroups have been considered; n is the number of individuals in each group; figure taken from paper III).

$CK/CK_{max}$ ) compared to their colleagues with only a basic introduction (Figure 4). No individual of these two groups with additional training assigned a wetness class which was off the group median by more than two classes. During the test in April the importance of additional training was not so evident among farmers as only the median  $CK/CK_{max}$  was higher for  $F_{trained}$  compared to  $F_{basic}$ , but not KA.

The qualitative soil wetness classes of the classification scheme for semi-arid conditions in Tanzania reflected differences in quantitative volumetric water content of the gravimetric soil samples taken during the test in April and June. The median volumetric water content increased with increasing wetness class but for the test in April classes 1, 2, 3, class 3 and 4 and classes 4, 5, 6, 7 were not significantly different from each other. For the dataset of the second test in June classes 1 and 2, classes 3, 4, 5 and classes 4, 5 and 6 were not significantly different from each other.

#### 4.2 Groundwater Variability

##### 4.2.1 Median Groundwater Levels

Groundwater dynamics varied spatially across the 20 ha mountain headwater catchment WS07 in the Alptal. Sites with a  $TWI < 4$  had predominantly positively skewed frequency distributions (i.e., mainly low water levels), while sites with a  $TWI > 6$  were predominantly negatively skewed (i.e., mainly high water levels). The skewness of the groundwater frequency distribution was correlated to all topographic indices considered in paper IV (e.g., local slope:  $r_s = 0.68$ ,  $TWI r_s = -0.69$ ), except local curvature. The fraction of time the wells were filled to a certain level below the soil surface was also related to topography.

The median relative groundwater levels were correlated to most of the selected topographic indices, however, the strength of the correlation differed for the local and upslope topographic characteristics. The median groundwater levels were correlated to the local slope ( $r_s = -0.67$ ) but not to the

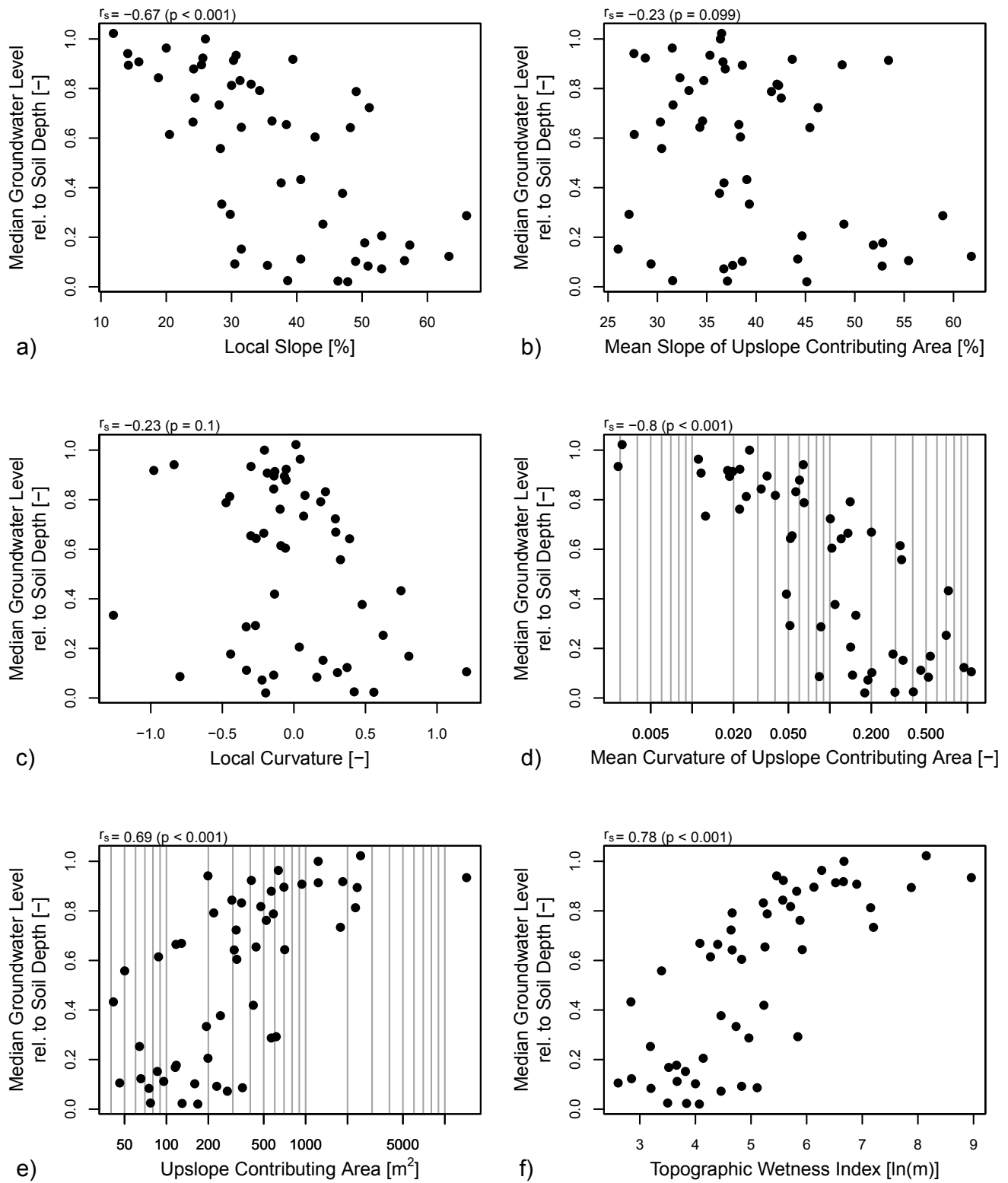


Figure 5: Median groundwater level relative to soil depth (1=at the soil surface, 0=at bottom of the well) as a function of local slope (a), mean slope of the upslope contributing area (b), local curvature (c), mean curvature of the upslope contributing area (d), upslope contributing area (e) and Topographic Wetness Index (f) (figure taken from paper IV).



mean slope of the upslope contributing area (Figure 5 a,b). In contrast, the median relative groundwater levels were highly correlated to the mean curvature of the upslope contributing area ( $r_s = -0.80$ ) but not to the local curvature (Figure 5c, d). The median relative groundwater levels were also correlated to the upslope contributing area ( $r_s = 0.69$ ), TWI ( $r_s = 0.78$ ) and the mean TWI of the upslope contributing area ( $r_s = 0.62$ ) (Figure 5e, f). Soil depth and the saturated hydraulic conductivity of the mineral soil were also considered to be important controls on median groundwater levels but the correlations were not statistically significant.

The correlation between TWI and absolute groundwater levels decreased strongly at the beginning of rainfall events and reached the lowest values shortly after peak streamflow. During the falling limb of the hydrograph,  $r_s$  increased quickly and reached the highest values twelve hours to two days after the event. During dry periods,  $r_s$  gradually decreased until the beginning of the next event. The drop in correlation at the beginning of a rainfall event was particularly large after long dry periods.

This event-scale change in correlation persisted throughout the year but was superimposed by a seasonal cycle:  $r_s$  was lowest during the dormant season ( $r_s$ : 0.5 to 0.6), intermediate during the growing season ( $r_s$ : 0.65 to 0.75) and highest during spring ( $r_s$ : 0.75 to 0.85). However, the highest  $r_s$  values were not necessarily tied to conditions of highest absolute streamflow or groundwater levels but rather to conditions of smallest relative change in streamflow and groundwater levels, respectively (Figure 6).

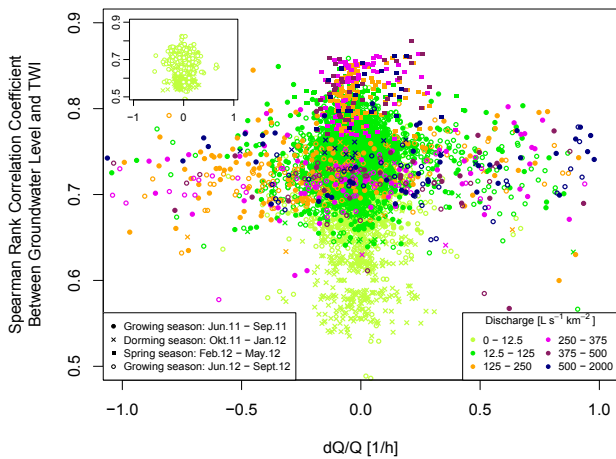


Figure 6: Spearman rank correlation coefficients ( $r_s$ ) between groundwater level and Topographic Wetness Index (TWI) plotted as a function of the relative change in streamflow at sub-catchment C5. Symbols represent different seasons, while colors represent different streamflow classes. The inset in the upper left corner shows the data for streamflow  $< 12.5 \text{ l s}^{-1} \text{ km}^{-2}$  without overlap of the other streamflow classes (taken from paper IV).

#### 4.2.2 Groundwater Response Timing

Half of the groundwater monitoring sites in the study catchment WS07 in the Alptal responded in 84% of all rainfall events and were delayed relative to the start of

rainfall by less than half an hour. Moderate-intense rainfall events had the shortest response times, which is in agreement with the perception that soil water deficits are satisfied faster when the rainfall intensity is higher. In fact,  $t_{\text{rise}}$  was highly correlated to  $P_{\text{rise}}$  ( $r_s = 0.98$ ). The inter quartile range (IQR) of the median  $t_{\text{rise}}$  of all sites was twice as high for the low-intense rainfall events than for the moderate-intense rainfall events but the difference in median  $t_{\text{rise}}$  for different antecedent conditions was not statistically significant.

The time to rise ( $t_{\text{rise}}$ ) was correlated to topographic indices, namely the mean curvature of the upslope contributing area ( $r_s = 0.82$ ), TWI ( $r_s = -0.81$ ), upslope contributing area ( $r_s = -0.74$ ), mean TWI of the upslope contributing area ( $r_s = -0.66$ ), and local slope ( $r_s = 0.64$ ). The  $r_s$  was lower for the mean slope of the upslope contributing area ( $r_s = 0.29$ ) and the local curvature ( $r_s = 0.28$ ). The median  $t_{\text{rise}}$  decreased with TWI for sites with a  $\text{TWI} < 6$  and was almost constant for all other sites. The slope of this decline of median  $t_{\text{rise}}$  with TWI was a function of rainfall intensity (Figure 7).

Groundwater peaks lagged rainfall centroids by about 1 hour during half of the rainfall events. Differences in  $t_{\text{peakP}}$  between the four rainfall event types were small, except for the low intensity events during dry antecedent conditions for which  $t_{\text{peakP}}$  was more than twice as long. Streamflow peaks were delayed relative to groundwater peaks by a few 10s of minutes but for 13 out of 51 sites, groundwater peaks lagged the discharge peak at the catchment outlet during half of all events. The timing of the groundwater peak (median  $t_{\text{peakP}}$  and  $t_{\text{peakQ}}$ ) was not correlated to any of the topographic indices when all events were considered.

The peak duration ( $t_{\text{dur}}$ ) was expected to be governed by subsurface inputs from upslope and therefore correlated to the upslope contributing area. Instead the median  $t_{\text{dur}}$  was found to be correlated only to local slope ( $r_s = -0.32$ ) and mean TWI of the upslope contributing area ( $r_s = 0.29$ ) but not to any of the other indices.

The median time to recession ( $t_{\text{rec}}$ ) was correlated to local slope ( $r_s = -0.39$ ) and TWI ( $r_s = 0.38$ ), as well as to the mean curvature of the upslope contributing area ( $r_s = -0.37$ ) and the upslope contributing area ( $r_s = 0.32$ ) but not to any of the other indices. The median  $t_{\text{rec}}$  increased with increasing TWI for sites with a  $\text{TWI} < 6$  but the variability was high. Differences between the four rainfall event types were small.

## 5. Discussion

### 5.1 Qualitative Soil Moisture Assessment

#### 5.1.1 Reliability of the Boots & Trousers Method

The qualitative Boots & Trousers method for soil wetness classification in humid environments captured the quantitative differences in volumetric water content. It was also shown that the method was robust, as the agreement of the wetness class assignments among the two groups of Swiss students in paper II was high. Still the degree of



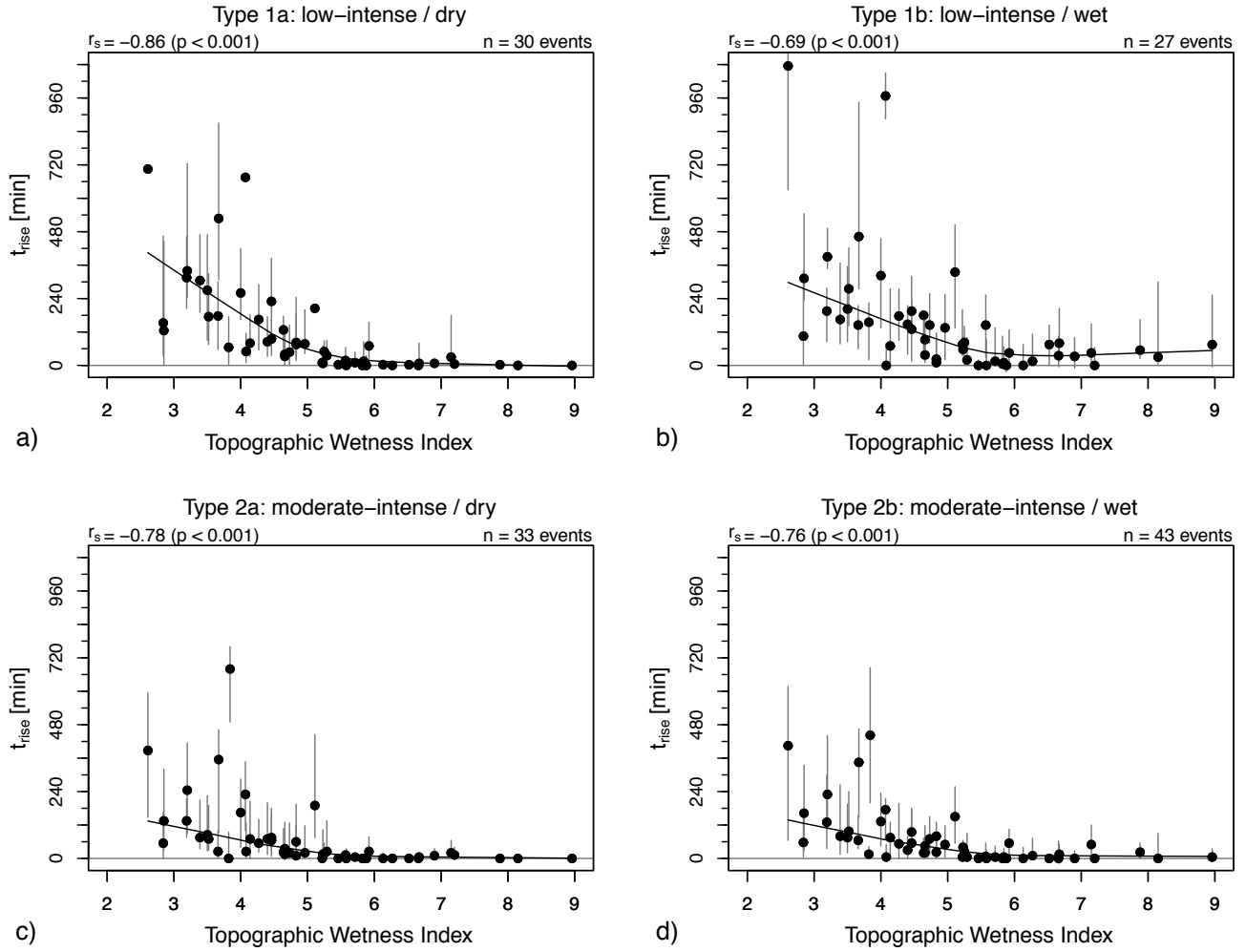


Figure 7: Time to rise ( $t_{rise}$ ) as a function of the Topographic Wetness Index for the four rainfall event types. Grey bar: inter quartile range, dot: median for each site, black line: LOWESS curves fitted to the median values,  $r_s$ : Spearman Rank Correlation Coefficient and associated p-value (figure taken from paper V).

agreement tended to be higher for group 1, which might be due to the fact that the sampling sites had dried up between the morning and the afternoon. Therefore more sampling points fell into the intermediate range of wetness, which seemed to be more difficult to classify. The degree of agreement was assumed to be even higher if raters would have had some previous experience or had gotten some training in how to identify the individual wetness classes. This was later proven to be correct in a similar test in Tanzania (paper III).

There is, however, potential to improve the method as far as differentiation of wetness classes is concerned. Especially sites, classified as dry had a considerable range in volumetric water content (dataset 2) and should be differentiated by more distinct indicators. On the other hand, the three intermediate and the two wettest classes could be combined, as the median and range in volumetric water content differed little from each other. Fewer wetness classes would likely be easier to assign and more distinct but fewer classes would lead to a coarser resolution of soil

moisture patterns and limit the characterization of changes in soil moisture over time. So it is suggested to use the seven wetness classes for mapping, leaving the option to later combine classes depending on the questions to be answered.

The Boots & Trousers method is fast to apply and needs no measurement equipment. Experience during field work showed that with this method one can assess about 5 times as many sampling points as with a mobile TDR device in the same time span. The new classification scheme can be applied without prior expert knowledge and has been shown to be robust and not significantly affected by rater subjectivity. All these advantages make this method particularly useful in remote areas and in developing countries (see paper III).

However, limitations exist as the classification method relies on wetness indicators that can be identified at the soil surface and therefore classifications are potentially influenced by vegetation or the litter layer. Drizzle, dew or evapotranspiration can also alter the appearance of the soil

surface without affecting the soil moisture at deeper depth. Regarding the absolute soil water content, the relation between qualitative wetness classes and quantitative soil moisture needs to be established for every soil type other than the Gleysol and Andosol soils tested in the two studies.

#### 5.1.2 The Effect of Experience and Training on the Inter-Rater Reliability

The qualitative soil wetness classification scheme for semi-arid environments that was tested in Tanzania proved to be a robust and intuitive method for mapping soil moisture differences. The agreement in wetness class assignments among  $S_{all}$  and  $E_{all}$  during the test in April and June was high. For  $F_{all}$  the agreement was lower during the first test but the within-group variability of class assignments was considerably reduced and gross misclassifications of up to 6 classes were avoided during the second test in June. A basic introduction in small subgroups, a redesign of the assessment form layout and clearer labeling of the sampling sites allowed  $F_{all}$  to agree or be within +/- one wetness class for more than 90% of all classifications. The dry to intermediate wetness classes seemed to be most difficult to assign, while the wettest classes were the easiest to assign. A profound basic introduction to the wetness classification scheme during the second test in June particularly improved the dry to intermediate class assignments by  $F_{all}$ . The benefit of a more detailed training was evident regardless of farming experience or education level for both  $F_{trained}$  and  $S_{trained}$ . Not only could the within-group agreement be improved but also the number of gross misclassifications was reduced.

The agreement of wetness classifications by  $S_{all}$  and  $E_{all}$  was similar to a test with master students in Switzerland (paper II). The agreement among  $F_{all}$  during the test in April was lower than in the Swiss study but reached a similar agreement during the second test with a better introduction. A better basic introduction also minimized the spread of class assignments and the bias of individual raters to classify wet sites as too dry and dry sites as too wet. While the mean classification difference of individual raters during the first test in April was much higher compared to the one in the study in Switzerland, it was similar during the second test in June.

The qualitative wetness classes reflected actual differences in volumetric water content, however the median values of the two driest classes and the three wettest classes were very similar suggesting that a classification scheme with fewer wetness classes would be sufficient to differentiate the actual range in volumetric water content. However as noted in paper II, a reduced number of classes would result in a coarser resolution of the resulting patterns and misclassifications would have a larger effect on the final result. It is also interesting to note that the median volumetric water content of class 1 in the Swiss study (38%) was similar to the median volumetric water content of class 7 (37%) in the Tanzanian study. This exemplifies that similar qualitative indicators on the soil

surface can be associated with different absolute volumetric water content and highlights the need to calibrate the scheme to the local soil type, if information about the absolute water content is needed.

Other limitations of this qualitative wetness classification scheme exist, as only the soil surface properties are assessed, but for many crops, the soil moisture at depth is of main interest. In principle the method could also be applied to a soil sample which is taken from a small pit, dug down to the rooting depth with a spade (Görbing & Sekera 1947). However soil moisture at the surface can be expected to be related to soil moisture at depth for most soil types if the vertical soil moisture profile is close to equilibrium. Soil wetness classifications directly after rainfall should thus be avoided. Other potentially influencing factors are the vegetation and litter, wetting by dew and drizzle and drying due to evaporation.

## 5.2 Groundwater Variability

### 5.2.1 Median Groundwater Levels and Topographic Controls

The correlation analysis in paper IV suggests that topography exerts a dominant control on the median groundwater levels in mountain catchments with low permeability soils. Median groundwater levels were related to local controls, such as the local slope and the soil wetness (as described by the TWI) and upslope controls, such as the runoff concentration within the upslope contributing area (as described by the mean upslope curvature), subsurface water input from upslope (as described by the upslope contributing area) and mean soil wetness in the source area (as described by the mean TWI of the upslope contributing area). Interestingly, the relative strength of slope and curvature in explaining the median groundwater levels depended on whether they were considered as local or upslope controls. Other studies also reported groundwater levels to be correlated to TWI, local slope and upslope contributing area, although the correlation coefficients were lower than in this study (Detty & McGuire 2010; Bachmair & Weiler 2012). Particularly on footslopes and in catchments with a relatively flat topography or conductive soils, the TWI was weakly correlated to the spatial variation in groundwater level (Moore & Thompson 1996; Seibert et al. 1997). This difference suggests that in steep mountain headwater catchments with low permeability soils, as in the Alptal, groundwater levels are predominantly shallow and strongly influenced by surface topography. In catchments with transmissive soils, saturation and subsequent lateral subsurface flow occurs deeper in the soil profile and therefore soil properties, such as the saturated hydraulic conductivity, soil depth, or subsurface topography, are expected to be of greater importance than in environments with low permeability soils (McDonnell 1990; Uchida et al. 2003; Tromp-van Meerveld & McDonnell 2006).

The variability in median groundwater levels was largest for sites in midslope locations with a local slope between 30% and 50%, an upslope contributing area between 200

and 600 m<sup>2</sup> and a TWI between 4 and 6. Flatter footslopes and steeper ridge sites were characterized by a smaller variability in median groundwater levels. It could be speculated that an interplay of several factors controls the median groundwater levels on the midslopes, while for the footslopes and ridges only a few factors are important. This makes prediction of median groundwater levels in footslope- and ridge sites more reliable than for midslopes and suggests that midslopes are most relevant in terms of monitoring changes in groundwater storage and hydrological connectivity.

The assumption of steady-state successions was best met during conditions of small changes in runoff (= near zero  $dQ/Q$ ) and presumably small changes in groundwater levels during the groundwater recession of single events and during the snow melt season. The TWI assumptions were, however, not fulfilled during large changes in groundwater levels and streamflow during the start of events, when spatial variability of rainfall inputs and subsurface flow from upslope areas, drainage and associated variability in groundwater responses were high. The saturated zone did also not respond in unison during the lowest flows at the end of long dry periods when some wells were dry and connectivity was likely lowest.

### 5.2.2 Groundwater Response Timing and Topographic Controls

The analysis in paper V revealed that timing of groundwater rise and recession was strongly controlled by topography in a catchment with low permeable soils and shallow groundwater tables. The way subsurface water flow is generated and concentrated in the upslope

contributing area and the local drainage conditions seem to determine the timing of the rise and recession of the groundwater levels during a rainfall event. The time to rise ( $t_{rise}$ ) was influenced more by topographic position than by rainfall characteristics. Ridges and backslopes had the highest soil water storage dynamics in the catchment and are therefore expected to be most sensitive to differences in rainfall event characteristics and antecedent wetness conditions (see also paper IV). At the same time, the analyses showed that the water level response on backslopes was delayed, while wet sites in footslope locations responded quickly and most likely dominated the rapid streamflow response on the rising limb of the hydrograph. The strong correlation between the  $t_{rise}$  and the sum of rainfall until  $t_{rise}$  ( $P_{rise}$ ) highlighted the role of the storage capacity on the time to rise. The strong correlation between  $P_{rise}$  and the topographic characteristics allows identification of the parts of a catchment that are likely to respond as a function of cumulative event precipitation (Figure 8). These patterns can help to understand the establishment of hydrologic connectivity during events.

In contrast, time to peak ( $t_{peakP}$  and  $t_{peakQ}$ ) was not correlated to the topographic indices but was more likely controlled by rainfall characteristics. The groundwater peaks preceded runoff peaks by less than 20 minutes or even lagged it. The latter is plausible and in agreement with other studies elsewhere that have shown based on end member mixing analysis of hydrochemicals and stable water isotopes that hillslopes mainly contribute to streamflow during the recession of streamflow (McGlynn & McDonnell 2003; Burns et al. 2001). The duration of the groundwater peak ( $t_{dur}$ ) was more dominated by local

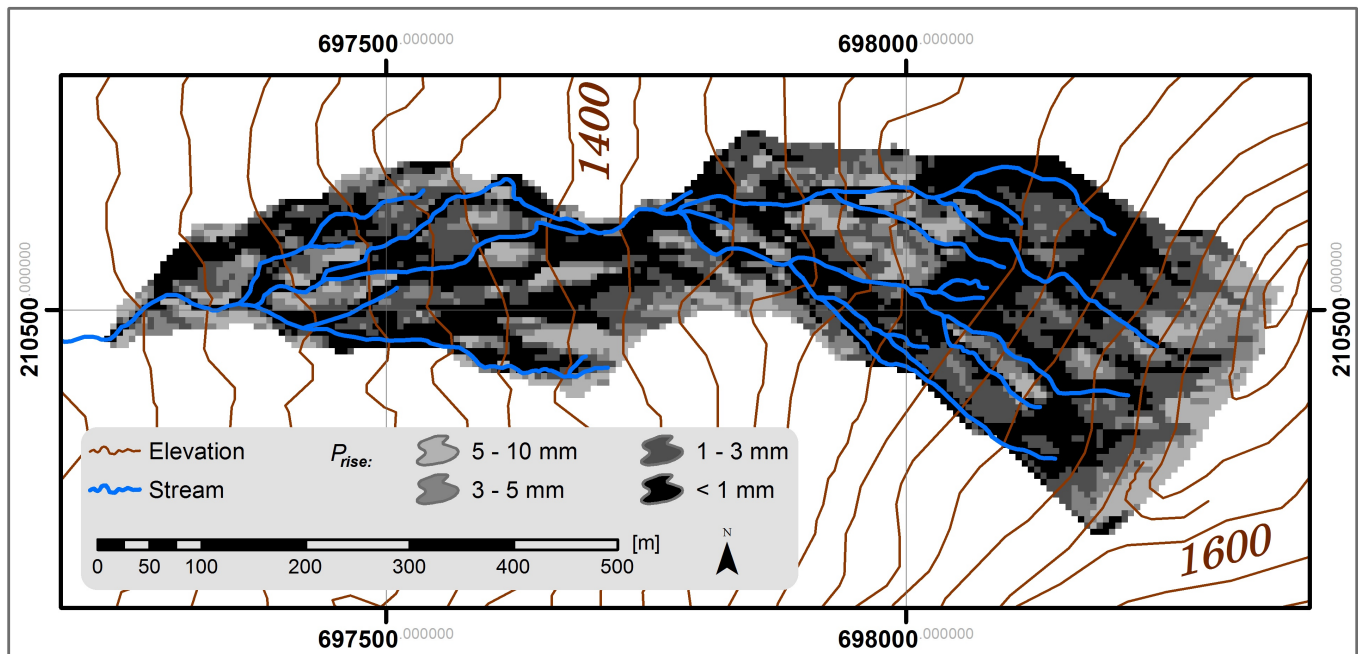


Figure 8: Example of the spatial distribution of an expected groundwater response after 1 mm, 3 mm, 5 mm and 10 mm of cumulative rainfall based on the relationship median  $P_{rise}$  and TWI. The aim of this thesis is to contribute to a better understanding of the underlying organizing principles that result in these spatial patterns (Background-topographic map: Swisstopo, 123456789) (figure taken from paper V).

drainage than by the subsurface contribution from the upslope. A possible explanation can be that the upslope contributing area was only partly hydrologically connected during events. The duration of the groundwater peak was influenced by rainfall intensity suggesting that only during events with higher intensity and potentially also larger event sum of precipitation, the upslope contributing area was more likely to be connected to the site, enabling a more persistent subsurface contribution. As expected, the antecedent wetness did not influence the timing and duration of the groundwater peak.

The duration of the recession ( $t_{rec}$ ) seemed to be controlled by the local drainage conditions and the water input from upslope. The groundwater recession ( $t_{rec}$ ) was longer and more variable for dry than wet antecedent conditions. However, this could be partly an artifact of the differences in the groundwater amplitude between dry and wet conditions. In fact, the slope of the recession ( $s_{rec}$ ) was not different for the four rainfall event types.

The results of paper V are seemingly in agreement with previous studies as they generally reported the fraction of wells, where a groundwater response was observed, to be highly correlated with total event precipitation, intermediately correlated with rainfall intensity and weakly or not correlated with antecedent conditions (Bachmair et al. 2012; Penna et al. 2014).

## 6. Conclusions

This thesis followed the idea of gaining better insight into catchment hydrologic functioning by analyzing hydrologic patterns, in this case groundwater- and soil moisture patterns, to decipher the organizing principles and dominant controls of hydrological responses at the catchment scale. The work focused on different kinds of patterns such as the spatial patterns in soil moisture for which classification schemes based on soft data were proposed in paper II and paper III. It also investigated patterns from hard data such as the groundwater measurements in paper IV and paper V. The motivation behind analyzing patterns was to infer dominant controls and organizing principles and thus to gain general understanding about physical processes of catchment hydrological functioning. These principles are expected to be transferable to similar catchments and even catchments of different scales. The research on patterns is driven by the general aim of being able to better predict groundwater, soil moisture and runoff response in headwater catchments for storms and base-flows, dry and wet environmental conditions and for current and future climate conditions.

This thesis is therefore not seen as yet another study contributing to the documentation of the idiosyncrasies of natural hydrological systems (McDonnell et al. 2007) but was designed to systematically learn from hypothesis testing about the hydrological behavior of a catchment with low permeability soils and the information inherent in qualitative soil moisture data. In the case of the soil wetness classification the test was on the hypothesis, that

qualitative indicators, which one can see, hear and feel, are distinct enough to result in robust wetness classifications when being applied by different raters (paper II & III). It was further hypothesized and tested if additional training results in better agreement among wetness classifications compared to only a basic introduction (paper III). In addition the hypothesis was tested that these qualitative wetness classes reflect actual differences in volumetric soil water content (paper II & III).

The main hypotheses underlying the analysis of groundwater variability was that different controls and runoff processes dominate in a catchment with low permeability soils compared to a catchment with transmissive soils (paper IV & V). More specifically, the TWI-assumptions were tested, of which the most important being that groundwater level variation can be approximated by a series of steady-state situations (paper IV). Another hypothesis tested in this thesis was on threshold-like groundwater level response governed by soil water deficits and therefore indirectly by topography (paper V). Better information on the dominant controls of groundwater dynamics was expected to shade some light on the question of why streams in the Alptal region respond so quickly and result in peak flows several orders of magnitude larger than their baseflow (Hegg et al. 2006).

### 6.1 Qualitative Soil Moisture Assessment

Paper II and paper III demonstrate the potential of soil wetness classification schemes based on qualitative indicators to capture shallow soil moisture differences in humid-temperate and semi-arid environments. It was important to adapt the qualitative indicators to the context of the local peoples' every-day life experience and soil type, which is why in paper III different indicators are suggested for the rural, semi-arid area in Tanzania than in paper II for the humid mountain area in Switzerland. For both schemes, the comparison of qualitative wetness classes to gravimetric and TDR measurements showed that the qualitative wetness classes reflect actual differences in volumetric water content of the soil with some overlap between individual classes. The results in paper II and paper III showed that neither experience nor a certain level of education are a prerequisite for robust wetness classifications, but the analysis in paper III highlighted the value of a detailed introduction and training to obtain a better agreement between individual raters. There was a high level of agreement between raters when classifying the wettest classes, while intermediate wetness classes seemed to be more difficult to assign. Some raters had a systematic tendency to rate specific wetness classes as too wet or too dry but when the raters were familiar with the application of the scheme, their mean offset was small and typically within the range of one wetness class.

A soil wetness classification scheme as presented in paper II and paper III is fast to apply, requires no experience and no measurement equipment, and still can provide robust and reliable data on soil moisture. The Boots & Trousers method and the adapted version for semi-arid

environments are seen as a supplement to existing quantitative measuring techniques and allow for quick determination of soil moisture patterns over a large area of interest. It could be shown that such qualitative methods can be applied successfully in a wider range of soil- and environmental conditions. All these advantages make the classification scheme particularly useful and appropriate for developing countries and remote areas.

## 6.2 Groundwater Variability and Dominant Controls

The objective of paper IV and paper V was to assess the dominant topographic controls on median groundwater levels and the importance of rainfall and antecedent conditions on the groundwater response timing in a 20 ha pre-alpine catchment with low permeability soils. Results of a rank correlation analysis with data from 51 groundwater monitoring sites and 133 rainfall events suggest that topography is a good predictor of the median groundwater level, the time to groundwater rise and the duration of the recession but not for the timing of the groundwater peak. Median groundwater levels were correlated with selected topographic indices calculated for the monitoring site and its upslope contributing area. This suggests that median groundwater levels were controlled both by local drainage and by subsurface inputs from upslope. Paper IV also showed that the rank correlation between groundwater levels and TWI was not constant over time but decreased at the beginning of rainfall events as groundwater levels responded differently throughout the catchment. After the event peak, when the catchment was slowly draining or during snowmelt in spring, the correlation between groundwater levels and TWI was highest. Under these conditions the TWI assumptions of a series of steady-state successions, connected upslope contributing areas and surface slope as a proxy of the hydraulic gradient were fulfilled best. These assumptions were least fulfilled during long dry periods, when parts of the catchment drained at different rates and became disconnected. This has implications for using TWI-based models to predict the spatial patterns of groundwater levels, their connectivity and the catchment runoff response.

A rainfall threshold for groundwater initiation existed, which was strongly dependent on topography. The relationships between TWI and the cumulative rainfall ( $P_{rise}$ ) and time to rise ( $t_{rise}$ ) allows spatial patterns of average groundwater response zones to be predicted (Figure 8). Rainfall intensity influenced the time to rise because it influenced the time needed to satisfy the soil moisture deficits. The antecedent wetness conditions in general turned out to play a minor role for the groundwater response timing in the study catchment as conditions were generally wet and groundwater levels were predominantly high.

Because the topographic indices were identified as good predictors of groundwater response timing in this study, while other studies suggested soil properties and bedrock topography to be more important, one can conclude that surface topography might play a more important role in

catchments with low permeability soils and predominantly shallow groundwater tables than in catchments with more transmissive soils (Tromp-van Meerveld & McDonnell 2006; Bachmair et al. 2012; Penna et al. 2014).

The dataset, comprising more than 50 monitoring sites and 133 rainfall events, allowed strong correlations to be revealed, especially between median groundwater levels and selected topographic indices and groundwater response timing and topographic predictors, rainfall characteristics and antecedent wetness. It also showed the very large variability in the response to different rainfall events. These relations might not have been clear for a smaller dataset.

## 7. Outlook

The work presented in this thesis offers multiple options for continued or improved research. The qualitative soil wetness classification scheme proved to be robust in two contrasting environments but the intermediate wetness classes had the largest spread in class assignments, which could be improved by a refinement of the class indicators. The idea of using blotting paper as stated in paper II was tested in Tanzania but not found to be helpful. As part of an ongoing master thesis it is being tested whether the Boots & Trousers method is also reliable if the method is only introduced on a handout with instructions but no person is explaining it or answering questions. The aim of this work is to identify the most efficient way to explain/introduce the method to a large number of users.

The qualitative soil wetness classification schemes have a wide range of applied and scientific applications. Soil moisture patterns can be informative for the potential distribution of plant habitats, for assessing of the optimal conditions for growing crops, the potential risk of crop failure if the plant available water is low or the risk of floods when the soil storage is almost full and additional rainfall is forecasted. Qualitative soil moisture patterns are useful for the calibration and validation of models and for data assimilation when using the models for predictions. Systematically testing the value of soft soil moisture data for model calibration and validation would potentially give it more attention than it currently has.

For the farmers in Tanzania the qualitative soil wetness classification scheme can be a tool for decision making to allocate the available water resources within a farming community in a fair way. The implementation of the qualitative soil wetness classification scheme in several farming villages in the study area near Arusha, Tanzania is the vision of the project “iMoMo - Innovative Monitoring and Modeling of Water” by the Swiss Agency for Development and Cooperation (SDC). One of the main objectives is to implement a crowd-sourced collection of distributed environmental data with thousands of sampling points. Trained farmers would send wetness classifications of their fields via SMS to a common decision support system and in return could be provided with predictions of soil water stress and suggestions on how to best use the available water resources.

Concerning the patterns of groundwater variability, this thesis covers the first steps of mapping/measuring patterns and identifying the dominant controls. In the course of a postdoc project called “ICaRuS – Investigating Catchment Runoff Response by Assessing Spatial Patterns of Groundwater Dynamics” that was granted to me by the Swiss National Science Foundation (SNF), it is intended to use the outcome of this thesis on dominant controls, to predict groundwater response at all non-monitored sites which will then result in maps of spatially distributed groundwater response patterns. The proposed strategies are:

- 1) Predicting zones of expected groundwater response as a function of cumulative event precipitation and the relation between median  $P_{rise}$ , median time to rise ( $t_{rise}$ ), and TWI described in paper V.
- 2) Clustering sites with similar groundwater hydrographs and site characteristics and using time series modeling to predict a representative groundwater response based on rainfall event characteristics. The representative groundwater hydrograph could then be assigned to all areas (pixels) that have similar site characteristics as the monitoring sites of each cluster.
- 3) Using a distributed model for simulating groundwater response patterns for individual rainfall events. The groundwater level measurements of paper IV and paper V could be used to calibrate/validate the model. A distributed groundwater model would also allow virtual experiments to assess the catchment's behavior under extreme conditions that are not included in the observation period (Weiler & McDonnell 2004).

In order to assess the quality of the spatial groundwater response patterns, the resulting maps need to be evaluated against independent measurements not used in the model calibration procedure.

The final step is to link spatial patterns of groundwater dynamics to the runoff response at the catchment outlet using connectivity statistics. Therefore the maps of spatially distributed groundwater level patterns need to be evaluated in terms of connections to the nearest stream or catchment outlet. The basic idea is that groundwater levels define the portion of the soil profile which is saturated and therefore potentially conducting water to neighboring sites and the stream network. The spatial patterns of groundwater levels can be analyzed in terms of:

- 1) the groundwater level threshold, rainfall amount and antecedent soil saturation that is needed to establish connectivity.
- 2) the landscape units that merge under different rainfall conditions.
- 3) the cluster size or percentage of saturated cells needed to cause a runoff response in the stream network.
- 4) the persistence of spatial groundwater response patterns after events or during the spring, summer, autumn or winter season.

- 5) a scale dependence or independence of the connectivity-discharge relation using data from the seven nested sub-catchments.

Outputs of this analysis will be functional relations between catchment-scale groundwater connectivity and the runoff response which are expected to be scale-independent and transferable to other catchments with similar environmental conditions.

## 8. Acknowledgments

I would like to highlight and express my gratitude to Jan Seibert, Ilja van Meerveld and Manfred Stähli for supervising my PhD thesis. They invested so much of their time in regular (Skype-) meetings, discussions and feedback on my analysis and writing which shaped the way I now approach scientific questions. Jan, as my main supervisor, gave me the freedom to develop and follow my own research interests but also guidance when I needed feedback on my work. Jan, I am especially grateful for your creative ideas, your advice on data analysis and your support and encouragement in situations, where I was struggling. Ilja offered to be my co-supervisor bringing in an enormous expertise in groundwater, subsurface flow processes and hydrological connectivity; the very topic I wanted to work on. She visited the field site with Jan and me and invited me for a research visit to the Vrije Universiteit Amsterdam in January 2013, which was the start of our ongoing collaboration. I am very grateful for your ideas and the feedback on my analysis and writing that considerably shaped and improved my scientific work and the joint publications. Manfred supported this PhD thesis from the beginning, when we had to get legal permission to install our instruments in the Alptal, which is part of a natural protection area. He provided me not only with valuable long-term data of the meteorological weather station run by WSL but especially with very constructive advice to better structure and focus the paper manuscripts. He was also a very supportive mentor in the period when I was applying for postdoc funding. I am thankful to Stephan Gruber who gave me valuable advice on data analysis and focus of the PhD in an early stage of the PhD before he left to Canada. I am glad he was on my committee bringing in a non-hydrologist's perspective. I am also very thankful to Russell Smith, who also gave me important guidance in an early stage of my PhD before also leaving to Canada. He went to the field with me several times giving me advice and direct support on how to build and install the v-notches and the groundwater wells and shared very valuable thoughts on the measuring concept and the PhD proposal. I highly appreciate the practical things I learned from you, Russell. Many thanks also to Laurent Pfister who was willing to act as the external examiner of my thesis and discuss scientific details during the defense. He has a wide expertise in various hydrological topics and has followed research on creative, novel measuring techniques such as using diatoms as natural tracers or thermal imagery for mapping riparian-streamflow connectivity.

Many thanks also to Benjamin Fischer, my PhD colleague, who was also working on rainfall-runoff processes in the Alptal. I am very happy having had you as a colleague to discuss ideas on how the Alptal catchments might function and am particularly thankful for your strong support out in the field and in the lab (isotope analysis). I remember you helping me to install the v-notch of subcatchment C3 on 2<sup>nd</sup> and 3<sup>rd</sup> of December 2010 with 20 cm of snow and temperatures below the freezing point. This is a collegial effort I will not forget! I also appreciate our open-minded collaboration with sharing data, ideas, feedback and a helping hand and an encouraging word whenever it was needed.

Along with Ben and Russell I also had a group of other colleagues who helped me in establishing the distributed groundwater and streamflow measuring network. Karl Steiner was willing to carry up the heavy wire-plates and the flumes of which one of them is only a 100 m below the Furggelenstock mountain peak (!) Kari, without your strong help, this would definitely not have been possible! I could count on the reliable support of Ellen Cerwinka, Jana Dusik, Seraina Kauer, Nadja Lavanga, Stephan Müller, Stefan Plötner, Paribesh Pradhan and Sandra Schärer during the field installations in 2010 and the test of the Boots & Trousers method in summer 2011 in the Alptal. Andrea Kollegger, applied the Boots & Trousers method during her MSc-thesis in summer and autumn 2010 in the Erlenbach collecting a dataset of thousands of TDR measurements and qualitative soil wetness assessments. Mirjam Zehnder conducted repeated slug & bail tests in 2012 for all 51 groundwater monitoring sites and Peter Herrmann maintained the entire monitoring network in 2013 and 2014 and installed two additional monitoring sites for a spring and a macro pore. Many thanks, Andrea, Mirjam and Peter, for your enormous effort and motivation during your MSc and the good time we had while working together. Many thanks also to Ivan Woodrich who taught me (a man whose all fingers) how to use a milling machine and Claudia Müller who taught me how to mix chemicals to produce standards for my ion chromatography analysis in the lab.

Many thanks also to the master students of course GEO471, spring term 2011, the students and staff of the Nelson Mandela University (Arusha) as well as the farmers of the village communities of Mungushi, Kichangani and Kigongoni for their participation in the field test of the Boots & Trousers method in the Alptal and Tanzania. Without your effort the test of the qualitative wetness classification schemes would not have been possible. I highly acknowledge the support of our local Tanzanian partners of the Pangani Basin Water Board, the Upper Kikuletwa Water Users Association (Mr. Tito Kitomari) and local village and furrow leaders. I thank Tobias Siegfried (iMoMo project coordinator), Hosea Sanga (local iMoMo-project manager), Pascal Oechslin, Beat Lüthi and Sebastian Stoll (field assistance), Matthieu Bolay (anthropo-technologic issues) and Alfayo Miseyeki (translation).

I would like to thank Dan Moore, Rick Hooper, Fabrizio Fenicia and Uwe Ehret for good discussion during their short visits/talks at the department. Many thanks also to Hans Stötter, my former supervisor of the Diploma thesis at Innsbruck University and Gertraud Meissl who supported me particularly at the start of my scientific career. Thanks to D. Heersink (Department of Mathematics, University of Zurich) for thoughts on the statistical analysis; T. Camek (IMKO GmbH) for TDR calibration issues. Many thanks also to family Heinzer and Annagreth Schuler & Werni Ruhstaller, wardens of the Furggelen- and Zwäckenhütte in the Alptal for their warm welcome and a ride with the snowmobile. Thanks to the Oberallmeindkorporation (OAK) Schwyz, municipality Alpthal and the Canton of Schwyz - Amt für Natur, Jagd und Fischerei, for cooperation.

I want to thank all my colleagues of the Hydrology & Climate (h2k) group, namely Nans Addor (sharing the office and R tips with me), Tracy Ewen (for collaboration on data analysis from the Rietholzbach and proofreading my manuscripts), Kristyna Falatkova (for so much fun during our social activities while you were visiting the group), David Finger (for thoughts on modeling), Daniel Furrer (for v-notch calibration at Irchel Park), Glenda Garcia-Santos (on her thoughts during lunch discussions), Michael Hilf (for “Hallo Michi” echos in the hallway), Michal Jenicek (for snowball fights in the Schächental), Seraina Kauer (who gave me so much support when coming to Zurich), Ully Koch (for bringing cake), Käthi Liechti (for nice skitouring), Ljubica Lindov (for a good atmosphere in the secretary office), Silviya Nikolova (for sharing the office with me and the gifts from the Black Sea), Hana Pevna (for wonderful origami), Sandra Pool (for support during the Boots & Trousers test with students at Irchel park and fun during lunch and coffee breaks), Bianca Rossmann (for the support from Austria), Sandra Röthlisberger (for lab support), Andrea Rücker (for sharing the office and good discussions over afternoon ice cream breaks, Philipp Schneider (for advise and help with all kinds of technical equipment), Patrizia Di Secli (for working out the finances), Ania Sikorska (being my office mate and hiking colleague), Maria Staudinger (sharing the office and R tips with me) Ludek Strouhal (for Czech beer tasting), Marc Vis (becoming a good hiking partner and for taking care of the tropical forest in our office, when I was abroad), Daniel Viviroli (for good discussions on applied and scientific hydrologic issues and becoming a good friend), Ling Wang (for social activities and wonderful art paintings).

Many Thanks also to my colleagues of the Geography Department. Especially to Samuel Abiven (for support on Lägern test site), Stefano Endrizzi (side project using the GEOTop model), Anett Hofmann (for soil surveying tips), Ruth Hunkeler (responsible for human resources at GIUZ), Philip Jörg (for support on satellite data), Damien Markulin (for GPS support), Ronald Schmidt (for support on geo data), Martin Steinmann (for poster plotting service), Volker Wichmann (SAGA-expert in Innsbruck), Guido Wiesenberger (for support on soil surveying).



I would especially like to highlight my thanks to my parents Christine and Christoph and my sister Susanne for their love and support as a family and their true interest in my scientific work. I remember well your visit to the Studibach catchment in the Alptal. Many thanks for your mental support and encouragement along challenges during this PhD. Many, many thanks also to my close friends in Austria, especially Tyrol, Switzerland, Italy, especially South Tyrol, Scotland, Hungary, Germany, Australia and Canada for your support, helping hand and encouraging words and for numerous Skype- and pub meetings. I am very happy our friendships continued despite the spatial distances and the difference in life priorities as many of you now have family. Many thanks also to all my new friends, who I met in Zurich, making me feel at home here!

I thank the University of Zurich (Kanton of Zurich) for institutional funding of my PhD position and the Swiss Agency for Development and Cooperation (SDC) for financial support of the project: “Qualitative Soil Moisture Assessment in Semi-arid Conditions (Tanzania/ Africa)” as part of the Global iMoMo Initiative ([www.imomohub.org](http://www.imomohub.org)). Thanks also to the Graduate School of the Geography Department for providing useful courses and travel funding. In particular I want to thank Ross Purves and Isabelle Gärtner-Roer.

## 9. References

- Ali, G.A. & Roy, A.G., 2009. Revisiting Hydrologic Sampling Strategies for an Accurate Assessment of Hydrologic Connectivity in Humid Temperate Systems. *Geography Compass*, 3(1), pp.350–374.
- Ali, G.A., Roy, A.G. & Legendre, P., 2010. Spatial relationships between soil moisture patterns and topographic variables at multiple scales in a humid temperate forested catchment. *Water Resources Research*, 46(10), pp.1–17.
- Anderson, A.E., Weiler, M., Alila, Y. & Hudson, R.O., 2010. Piezometric response in zones of a watershed with lateral preferential flow as a first-order control on subsurface flow. *Hydrological Processes*, 24(16), pp.2237–2247.
- Anderson, M.G. & Burt, T.P., 1978. The role of topography in controlling throughflow generation. *Earth Surface Processes and Landforms*, 3, pp.331–334.
- Bachmair, S. & Weiler, M., 2012. Hillslope characteristics as controls of subsurface flow variability. *Hydrology and Earth System Sciences*, 16(10), pp.3699–3715.
- Bachmair, S., Weiler, M. & Troch, P. a., 2012. Intercomparing hillslope hydrological dynamics: Spatio-temporal variability and vegetation cover effects. *Water Resources Research*, 48(5), p.W05537.
- Barling, R.D., Corporation, W., Moore, I.D. & Grayson, B., 1994. A quasi dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content. *Water Resources*, 30(4), pp.1029–1044.
- Beven, K.J. & Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1), pp.43–69.
- Blazkova, S., Beven, K.J. & Kulasova, A., 2002. On constraining TOPMODEL hydrograph simulations using partial saturated area information. *Hydrological Processes*, 16(2), pp.441–458.
- Blume, T., Zehe, E. & Bronstert, A., 2009. Use of soil moisture dynamics and patterns at different spatio-temporal scales for the investigation of subsurface flow processes. *Hydrology and Earth System Sciences*, 13, pp.1215–1234.
- Boorman, D.B., Hollist, J.M. & Lilly, A., 1995. *Hydrology of soil types: a hydrologically based classification of the soils of the United Kingdom*, Institute of Hydrology, Wallingford, p. 146.
- Burns, D. a., McDonnell, J.J., Hooper, R.P., Peters, N.E., Freer, J.E., Kendall, C. & Beven, K., 2001. Quantifying contributions to storm runoff through end-member mixing analysis and hydrologic measurements at the Panola Mountain Research Watershed (Georgia, USA). *Hydrological Processes*, 15(10), pp.1903–1924.
- Burt, T.P. & Butcher, D.P., 1985. Topographic controls of soil moisture distributions. *Journal of Soil Science*, 36(1978), pp.469–486.
- Chaturvedi, R., Raghubanshi, A. & Singh, S., 2011. Plant functional traits with particular reference to tropical deciduous forests : A review. *Journal of Biosciences*, 36(5), pp.963–981.
- Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, pp.37–46.
- Darwin, C., 1859. *On the origins of species by means of natural selection* 1st ed., John Murray, London, p. 247.
- Detty, J.M. & McGuire, K.J., 2010. Topographic controls on shallow groundwater dynamics: implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment. *Hydrological Processes*, 24(16), pp.2222–2236.
- Dhakal, A.S. & Sullivan, K., 2014. Shallow groundwater response to rainfall on a forested headwater catchment in northern coastal California: implications of topography, rainfall, and throughfall intensities on peak pressure head generation. *Hydrological Processes*, 28, pp.446–463.
- Dooge, J.C.I., 1986. Looking for Hydrologic Laws. *Water Resources Research*, 22(9), pp.46–58.
- Dunne, T. & Black, R.D., 1970. Partial Area Contributions to Storm Runoff in a Small New England Watershed. *Water Resources Research*, 6(5), pp.1296 – 1311.
- Dunne, T., Moore, T.R. & Taylor, C.H., 1975. Recognition and prediction of runoff producing zones in humid regions. *Hydrological Sciences Bulletin*, 20(3), pp.305–327.
- Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. & D, P., 1991. Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta geobotanica*, 18, p.248.
- Evans, S., 1980. An integrated system of terrain analysis and slope mapping. *Zeitschrift für Geomorphologie*, Supl. Bd.(36), pp.274–294.
- Fannin, R.J., Jaakkola, J., Wilkinson, J.M.T. & Hetherington, E.D., 2000. Hydrologic response of soils to precipitation at Carnation Creek, British Columbia, Canada. *Water Resources Research*, 36(6), pp.1481–1494.
- Feyen, H., Wunderli, H., Wydler, H. & Papritz, A., 1999. A tracer experiment to study flow paths of water in a forest soil. *Journal of Hydrology*, 225(3–4), pp.155–167.
- Fischer, B., Rinderer, M., Schneider, P., Ewen, T. & Seibert, J., Spatial hydrochemical and isotope variability in six pre-alpine headwaters. *Hydrological Processes (in review)*.
- Gillin, C., Bailey, S., McGuire, K. & Gannon, J., 2014. Mapping of hypedopedologic spatial patterns in a steep headwater catchment. *Soil Science Society of America Journal*, (online), pp.1–51.
- Görbing, J. & Sekera, F., 1947. *Die Spatendiagnose - Ziel und Grundlage der zweckmässigen Bodenbearbeitung*, Hannover.
- Grayson, R. & Blöschl, G., 2000. *Spatial patterns in catchment hydrology: observations and modelling*, Cambridge University Press, Cambridge.
- Grayson, R.B., Western, A.W., Chiew, F.H.S. & Blöschl, G., 1997. Preferred states in spatial soil moisture patterns: Local and nonlocal controls. *Water Resources*, 33(12), pp.2897–2908.
- Güntner, A., Seibert, J. & Uhlenbrook, S., 2004. Modeling spatial patterns of saturated areas: An evaluation of different terrain indices. *Water Resources Research*, 40(5), pp.1–19.
- Haugt, D.R.W. & van Meerveld, H.J., 2011. Spatial variation in transient water table responses: differences between an upper and lower hillslope zone. *Hydrological Processes*, 25(25), pp.3866–3877.
- Hegg, C., McArdell, B.W. & Badoux, A., 2006. One hundred years of mountain hydrology in Switzerland by the WSL. *Hydrological Processes*, 20(2), pp.371–376.
- Hewlett, J. & Hibbert, A., 1967. Factors affecting the response of small watersheds to precipitation in humid areas. In S. W.E. & L. H.W., eds. *Proceedings of first International Symposium on Forest Hydrology*. Pergamon Press, pp. 275–290.
- Kollegger, A., 2010. *Untersuchung der räumlichen und zeitlichen Verteilung der Bodenfeuchtigkeit am Beispiel des Erlenbaches (Alptal/Schweiz). Vergleich einer quantitativen und einer qualitativen Messmethode (in German)*. Master thesis at the Department of Geography, University of Zurich. p. 71.



- Komakech, H.C. & Van der Zaag, P., 2011. Understanding the Emergence and Functioning of River Committees in a Catchment of the Pangani Basin, Tanzania. *Water Alternatives*, 4(2), pp.197–222.
- Krippendorff, K., 2004. *Content Analysis, an Introduction to its Methodology* 2nd ed., Sage Publications, Thousand Oaks, p. 440.
- Kulasova, A., Beven, K.J., Blazkova, S.D., Rezacova, D. & Cajthaml, J., 2014. Comparison of saturated areas mapping methods in the Jizera Mountains, Czech Republic. *Journal of Hydrology and Hydromechanics*, 62(2), pp.160–168.
- Lana-Renault, N., Regüés, D., Serrano, P. & Latron, J., 2014. Spatial and temporal variability of groundwater dynamics in a sub-Mediterranean mountain catchment. *Hydrological Processes*, 28(8), pp.3288–3299.
- Lin, H., Brooks, E., McDaniel, P. & Boll, J., 2008. Hypedology and surface/subsurface runoff processes. In M. G. Anderson, ed. *Encyclopedia of Hydrological Sciences*. Wiley Online Library, pp. 1–25.
- McDonnell, J.J., 1990. A Rationale for Old Water Discharge through Macropores in a Steep, Humid Catchment. *Water Resources Research*, 26(11), pp.2821–2832.
- McDonnell, J.J., Sivapalan, M., Vache, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M.L., Selker, J. & Weiler, M., 2007. Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, 43(7), pp.1–6.
- McGlynn, B.L. & McDonnell, J.J., 2003. Quantifying the relative contributions of riparian and hillslope zones to catchment runoff. *Water Resources Research*, 39(11), pp.1310–1320.
- McNamara, J.P., Chandler, D., Seyfried, M. & Achet, S., 2005. Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment. *Hydrological Processes*, 19(20), pp.4023–4038.
- Metcalfe-Smith, J.L., 1994. Biological Water-Quality Assessment of Rivers: Use of Macroinvertebrate Communities. In P. Calow & G. E. Petts, eds. *The River Handbook*. Blackwell Science Ltd., Oxford, pp. 144–171.
- Moore, R.D. & Thompson, J.C., 1996. Are Water Table Variations in a Shallow Forest Soil Consistent with the TOPMODEL Concept? *Water Resources Research*, 32(3), pp.663–669.
- Mosley, M.R., 1979. Streamflow generation in a forested watershed, New Zealand. *Water Resources Research*, 15(4), pp.795–806.
- Natural Resources Conservation Service, 1998. Estimating Soil Moisture by Feel and Appearance. *Program Aid*, 1619, pp.1–7.
- Penna, D., Mantese, N., Hopp, L., Borga, M. & Dalla Fontana, G., 2014. Spatio-temporal variability of piezometric response on two steep alpine hillslopes. *Hydrological Processes*, 29(2), pp.198–211.
- Peschke, G., Etzenberg, C., Müller, G., Töpfer, J. & Zimmermann, S., 1999. Das wissenschaftliche System FLAB - ein Instrument zur rechnergestützten Bestimmung von Landschaftseinheiten mit gleicher Abflussbildung. *IHI Schriften*, (10), p.122.
- Pfister, L., McDonnell, J.J., Hissler, C. & Hoffmann, L., 2010. Ground-based thermal imagery as a simple, practical tool for mapping saturated area connectivity and dynamics. *Hydrological Processes*, 3132(May), pp.3123–3132.
- Pielmeier, C. & Schneebeli, M., 2003. Developments in the Stratigraphy of Snow. *Surveys in Geophysics*, 24(5/6), pp.389–416.
- De Quervain, M., 1950. Die Festigkeitseigenschaften der Schneedecke und ihre Messung. *Geofisica Pura e Applicata*, 18, pp.179–191.
- Quinn, P., Ostendorf, B., Beven, K. & Tenhunen, J., 1998. Spatial and temporal predictions of soil moisture patterns and evaporative losses using TOPMODEL and the GASFLUX model for an Alaskan catchment. *Hydrology and Earth System Sciences*, 2(1), pp.51–64.
- Rodhe, A. & Seibert, J., 2011. Groundwater dynamics in a till hillslope: flow directions, gradients and delay. *Hydrological Processes*, 25(12), pp.1899–1909.
- Scherrer, S. & Naef, F., 2003. A decision scheme to indicate dominant hydrological flow processes on temperate grassland. *Hydrological Processes*, 17(2), pp.391–401.
- Schleppi, P., Muller, N., Feyen, H., Papritz, A., Bucher, J.B. & Flüher, H., 1998. Nitrogen budgets of two small experimental forested catchments at Alptal, Switzerland. *Forest Ecology and Management*, 101, pp.177–185.
- Schmocker-Fackel, P. & Naef, F., 2010. Changes in flood frequencies in Switzerland since 1500. *Hydrology and Earth System Sciences Discussion*, 7(1), pp.529–560.
- Schmocker-Fackel, P., Naef, F. & Scherrer, S., 2007. Identifying runoff processes on the plot and catchment scale. *Hydrology and Earth System Sciences*, 11(2), pp.891–906.
- Schüler, G., 2005. Herleitung von abflussrelevanten Flächen zur Steuerung von Wasserrückhaltmassnahmen im Wald. *Freiburger Forstliche Forschung*, (62), pp.143–158.
- Seibert, J., Bishop, K. & Nyberg, L., 1997. A test of TOPMODEL's ability to predict spatially distributed groundwater levels. *Hydrological Processes*, 11, pp.1131–1144.
- Seibert, J., Bishop, K., Rodhe, A. & McDonnell, J.J., 2003. Groundwater dynamics along a hillslope: A test of the steady state hypothesis. *Water Resources Research*, 39(1), pp.1014–1023.
- Seibert, J. & McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, 38(11), pp.1–14.
- Sidle, R.C., Tsuboyama, Y., Noguchi, S., Hosoda, I., Fujieda, M. & Shimizu, T., 2000. Stormflow generation in steep forested headwaters: a linked hydrogeomorphic paradigm. *Hydrological Processes*, 14(3), pp.369–385.
- Sim, J. & Wright, C., 2005. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), pp.257–268.
- Sivapalan, M., 2005. Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale. In M. G. Anderson & J. J. McDonnell, eds. *Encyclopedia of Hydrological Sciences*. Wiley, Chichester, pp. 193–219.
- Sklash, M. & Farvolden, R., 1979. The role of groundwater in storm runoff. *Journal of Hydrology*, 43, pp.45–65.
- Sklash, M., Stewart, M. & Pearce, A., 1986. Storm runoff generation in humid headwater catchments 2. A case study of hillslope and low-order stream response. *Water Resources Research*, 22, pp.1273–1282.
- Soil Survey Division Staff, 1993. *Soil survey manual*, United States Department of Agriculture - Soil Conservation Service, Washington D.C., p. 315.
- Soil Survey Staff, 2010. *Field Indicators of Hydric Soils in the United States. A Guide for Identifying and Delineating Hydric Soils* Version 7., United States Department of Agriculture - Natural Resources Conservation Service, Washington D.C., p. 53.
- Soulsby, C. & Tetzlaff, D., 2008. Towards simple approaches for mean residence time estimation in ungauged basins using tracers and soil distributions. *Journal of Hydrology*, 363(1–4), pp.60–74.
- Spearman, C., 1904. The proof and measurement of association between two things. *Journal of Psychology*, 15, pp.72–101.
- Tarboton, G.D., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources*, 33(2), pp.309–319.
- Tetzlaff, D., McDonnell, J.J., Uhlenbrook, S., McGuire, K.J., Bogaart, P.W., Naef, F., Baird, A.J., Dunn, S.M. & Soulsby, C., 2008. Conceptualizing catchment processes: simply too complex? *Hydrological Processes*, 22(11), pp.1727–1730.
- Thien, S.J., 1979. A flow diagram for teaching texture by feel analysis. *Journal of Agronomic Education*, 8, pp.54–55.
- Tilch, N., Zillgens, B., Uhlenbrook, S., Leibundgut, C., Kirnbauer, R. & Merz, B., 2006. GIS-gestützte Ausweisung von hydrologischen Umsatzräumen und Prozessen im Löhnersbach-Einzugsgebiet (Nördliche Grauwackenzone, Salzburger Land). *Österreichische Wasser-und Abfallwirtschaft*, 58(9), pp.141–151.
- Travis, M.R., Iverson, W.D., Gary, H. & Johnson, C.G., 1975. *VIEWIT: computation of seenarea, slope and aspect for land-use planning PSW 11*, USDA - Forest Service, Berkeley, California U.S.A., p. 70.
- Troch, P.A., Mancini, M., Paniconi, C. & Wood, E.F., 1993. Evaluation of a Distributed Catchment Scale Water Balance Model. *Water Resources*, 29(6), pp.1805–1817.
- Tromp-van Meerveld, H.J. & McDonnell, J.J., 2006. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, 42(2), pp.1–11.
- Uchida, T., Asano, Y., Ohte, N. & Mizuyama, T., 2003. Seepage area and rate of bedrock groundwater discharge at a granitic unchanneled hillslope. *Water Resources Research*, 39(1), p.1018.
- Uhlenbrook, S., 2003. An empirical approach for delineating spatial units with the same dominating runoff generation processes. *Physics and Chemistry of the Earth*, 28(6–7), pp.297–303.

- Violle, C., Navas, M.L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I. & Garnier, E., 2007. Let the concept of trait be functional! *Oikos*, 116, pp.882–892.
- Waldenmeyer, G., 2003. Abflussbildung und Regionalisierung in einem forstlich genutzten Einzugsgebiet (Dürreychtal, Nordschwarzwald). *Karlsruher Schriften zur Geographie und Geoökologie*, (20), p.192.
- Weiler, M. & McDonnell, J., 2004. Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology. *Journal of Hydrology*, 285(1-4), pp.3–18.
- Western, A. & Grayson, R., 2000. Soil Moisture and Runoff Processes at Tarrawarra. In R. Grayson & G. Blöschl, eds. *Spatial Patterns in Catchment Hydrology - Observations and Modelling*. Cambridge University Press, Cambridge, pp. 209–246.
- Western, A.W., Blöschl, G. & Grayson, R.B., 2001. Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources*, 37(1), pp.83–97.
- Whipkey, R.Z., 1965. Subsurface stormflow from forested slopes. *Bulletin of the International Association of Scientific Hydrology*, 10(2), pp.74 – 85.

# APPENDIX



# PAPER I



# Soil Information in Hydrologic Models: Hard Data, Soft Data, and the Dialog between Experimentalists and Modelers

Michael Rinderer<sup>1,\*</sup> and Jan Seibert<sup>1,2</sup>

## ABSTRACT

For understanding and predicting rainfall–runoff processes in watersheds, soils and their hydraulic properties play a central role. Experimentalists observe and document hydric soil indicators in detail for more and more sites in various catchments. Modelers, on the other hand, try to break down natural process complexity into models that are based on simplified process descriptions. The challenge for both is to identify first-order controls of catchment hydrologic behavior, which helps to better understand the nonlinearity of natural systems. This chapter describes how both, experimentalists and modelers, can work together toward a better understanding and quantification of subsurface runoff processes. Specifically, this chapter addresses the following questions: (1) How are subsurface runoff processes represented in models of different complexity, ranging from simple conceptual ones to more complex physically based ones? (2) How can catchment-scale models be parametrized using point-scale measurements and existing model approaches originally developed for small scales (e.g. a soil column)? (3) Which information can be gained from soil surveying methods, including mapping approaches of hydric soil indicators? (4) Can decision schemes be useful to indicate dominant runoff processes in an objective way? Finally we describe the soft data concept as a possible way forward to enhance the dialog between experimentalists and modelers. Soft data refer to all kinds of qualitative or semi-quantitative information on pedologic and hydrologic processes and properties. These data can be made useful for modeling by applying fuzzy-logic-based functions to evaluate the degree of acceptance of model simulation outputs compared to experimentalists' field experience.

<sup>1</sup>Department of Geography, University of Zurich, Winterthurerstr 190, CH-8057 Zurich, Switzerland

<sup>2</sup>Department of Earth Sciences, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden

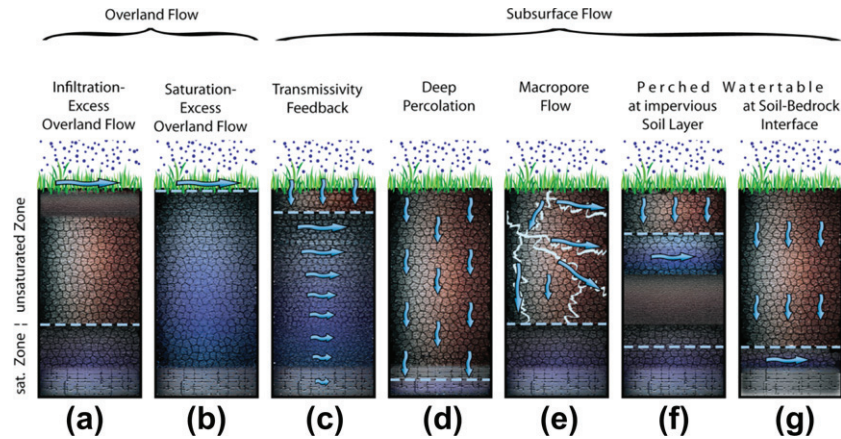
\*Corresponding author: Email: michael.rinderer@geo.uzh.ch

## 1. DIFFERENT VIEWS ON SOIL HYDROLOGIC FUNCTION – AN INTRODUCTION

Experimentalists who have gained experience on rainfall–runoff processes during their fieldwork can confirm that the governing processes are complex and often show a nonlinear, threshold-type behavior, especially when considering the scale of an entire catchment (Tromp-van Meerveld and McDonnell, 2006). While experimentalists tend to document natural phenomena in detail, current models are often a gross simplification of this natural complexity. In other words, complex processes are expressed in the form of mathematical equations based on simplifying assumptions such as a steady-state groundwater response throughout the entire catchment or flow pathways determined by topography. Often these assumptions contradict with what is observed in the field. Improved discussions between experimentalists and modelers (Seibert and McDonnell, 2002; Weiler and McDonnell, 2004) have shown potential to improve process understanding and correct modeling for the right reasons (Kirchner, 2006; Klemes, 1986). Existing process conceptualizations being at odds with hydrometric, hydrochemical, and isotope measurements have forced hydrologists to reassess and extend existing process conceptualization and find new, more appropriate explanations of runoff generation (McGlynn et al., 2002). In this chapter, we focus on the potential of a dialog between experimentalists contributing their wealth of somehow intuitive field experience – called tacit knowledge – and the modelers, who are strong in conceptualizing and therefore structuring processes to distill natural complexity down to its first-order controls of runoff generation in catchments. We limit our discussion to subsurface runoff processes, which we define as saturated water flow phenomenon in the soil of a catchment either due to the rise of an existing water table into more transmissive soil layers (Fig. 1c) or due to the transient saturation above an impeding layer, soil–bedrock interface, or some zone of reduced permeability at depth (Fig. 1f,g), both causing a subsequent lateral flow component and often a more or less delayed response of stream flow discharge (Weiler et al., 2005). Included in our discussion are runoff processes due to exceeding infiltration capacity or saturation at the soil surface (Fig. 1a,b). The term soil water flow is used in circumstances when the process of redistribution of infiltrated rainfall in a soil profile through the soil matrix and/or through preferential flow pathways is meant. We discuss the potential of semi-quantitative or qualitative data (called soft data) and their value in model calibration and model structure optimization, especially in catchments with little or no data availability. In particular, we address the following questions:

- 1) How are soils represented in different types of hydrologic models? What kind of conceptual and physically based approaches do exist to simulate key processes of subsurface runoff?
- 2) How can soil routines used at the catchment scale be parametrized? What is the value and limitation of point measurements made at the plot scale and





**FIGURE 1** Different types of surface and subsurface runoff processes: If infiltration capacity (a) or storage capacity (b) of the soil layers at or near the surface is exceeded, overland flow is generated. As groundwater rises into more transmissive soil layers (c), lateral subsurface runoff increases. If the soil has a coarse texture and the bedrock is highly permeable, the groundwater table tends to be low and soil water can percolate deep down into the soil or bedrock (d). If distinct horizontal macropores exist, the flow rates toward the nearest stream channel can be high and subsurface runoff therefore responsive (e). If a water-restricting layer exists within the soil profile (f) or at the soil–bedrock interface (g), a perched water table might form from time to time and lateral subsurface flow can occur within the saturated soil layer. (Color version online and in color plate)

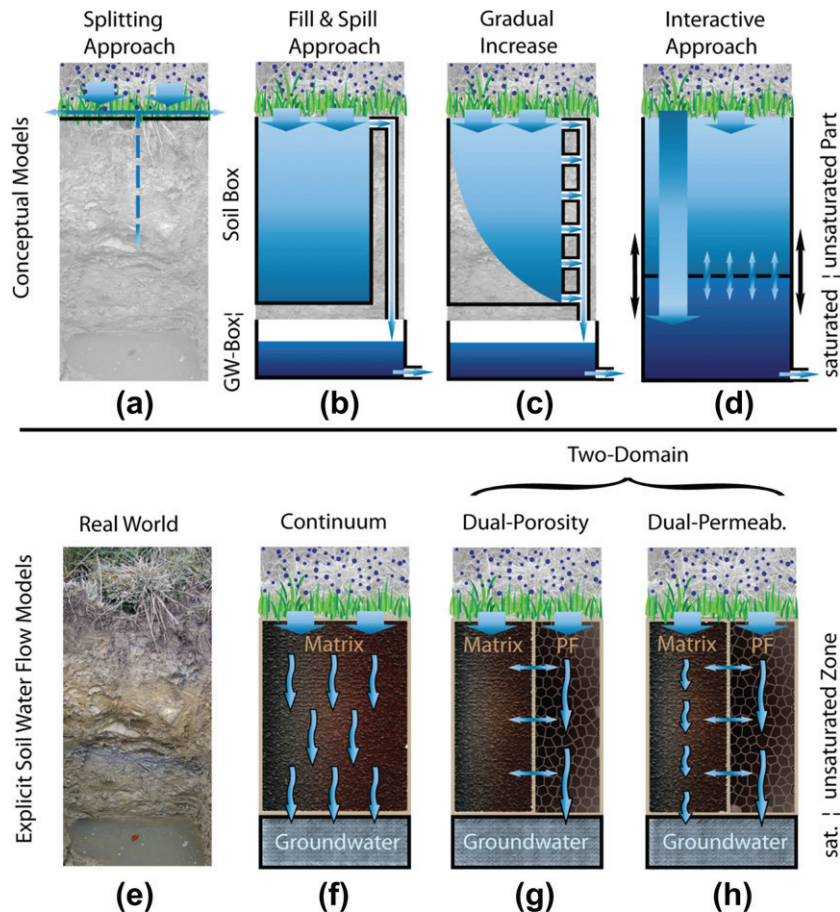
what kind of problems arise when upscaling this information to the catchment scale?

- 3) What kind of tacit knowledge as well as hard and soft indicators do already exist in soil survey data sets?
- 4) How can tacit knowledge and hard and soft indicators be linked in a systematic and objective way to indicate subsurface runoff processes?
- 5) How can the dialog between experimentalists and modelers be enhanced by utilizing soft data and tacit knowledge?

## 2. SOIL AND SOIL PARAMETERS IN HYDROLOGIC MODELS – THE MODELERS' POINT OF VIEW

### 2.1. Conceptualization of Subsurface Runoff Processes

This section briefly assembles and discusses the main modeling concepts currently used in conceptual and physically based hydrologic models for simulating subsurface runoff and soil water flow at the plot to catchment scale. This overview is not exhaustive but intends to raise the readers' awareness of the sometimes very simplistic view of modelers on subsurface runoff processes compared to the experimentalists' view, which focuses on the complexity of natural processes in all details (see Section 3).



**FIGURE 2** Conceptual and physically based modeling approaches of subsurface flow or soil storage response: Upper row shows conceptual modeling approaches of soil-water-storage outflow: (a) only accounting for surface runoff caused by infiltration excess; generating soil water outflow if bucket is (b) completely or (c) partly filled; and accounting for (d) interactions between saturated and unsaturated soil water storage. Lower row shows explicit soil water flow modeling approaches which can be categorized as (f) continuum approaches, which only account for matrix flow, and (g, h) two-domain approaches, which consider both matrix and preferential flow (PF). (Color version online and in color plate)

In very simple rainfall–runoff models, which only cover infiltration-excess overland flow (Figs. 1a and 2a), subsurface runoff is not simulated explicitly. The infiltration capacity of the soil surface determines the portion of rainfall, directly contributing to overland flow, which is of primary interest in many applied modeling cases to estimate flood discharge in small catchments. The simplest example of this model type is the rational method developed by [Mulaney \(1851\)](#) or the unit hydrograph models based on [Sherman \(1932\)](#), which are still in use to estimate peak discharge (e.g. [Hua et al., 2003](#); [Rinderer et al., 2009](#)).

Most hydrologic models do account explicitly for subsurface runoff processes. Conceptual models often mimic the property of the soil to store and release infiltrated water by a storage component with a defined storage–outflow relationship. The capacity for total water storage within a soil is dependent on its soil porosity and soil depth, but in most conceptual models only the dynamic storage is considered, which depends on drainable porosity. The soil storage must either first fill up to its maximum capacity before outflow commences (i.e. fill-and-spill) (Fig. 2b) or gradually increase its outflow (Fig. 2c). Stored water can be lost due to evaporation, to groundwater recharge, or to exfiltration. The latter is not represented in most conceptual models. Actual evaporation is usually estimated as a function of potential evaporation and the degree of soil saturation. Groundwater recharge can be simulated in a similar manner based on the amount of infiltrated water and the current soil-water-storage content. To account for interaction between saturated and unsaturated storage, the total soil water storage can be divided into two compartments with the boundary between them moving up or down according to the water budget in the storage (Fig. 2d) (Seibert et al., 2003; Seibert et al., 2011).

In more complex models, not only the net soil water outflow from a bucket is computed, but also the vertical and/or lateral flux of water between soil units is simulated explicitly. Under the consideration of soil as a saturated, porous media with homogeneous properties, the velocity of flow through this media can be approximated by Darcy's law. Richards (1931) generalized Darcy's law to be applicable for unsaturated soil conditions by assuming the same linear relationship, but with the hydraulic conductivity varying nonlinearly according to the degree of soil saturation. Richards further combined the Darcy–Buckingham law with the continuity equation to result in a flow equation known as the Richards equation. To avoid computationally intense calculations – related to solving the Richards equation – the kinematic wave approach can be used as an approximation, which combines the continuity equation with a storage–flow relationship assuming a functional dependency between them.

Under natural conditions additional to matrix flow (Fig. 2f), often other types of soil water flow are observed. Therefore, more enhanced models – so-called two- or multidomain models (Köhne et al., 2009) – distinguish between rapid macroporous flow (e.g. flow through structured pores, cracks and fissures, and wormholes) and slow seepage through the soil matrix (Fig. 2g,h). Transfer terms account for the exchange between the flow domains. Domains are a conceptualization rather than a geometric separation of the soil volume. Two types of two- or multidomain models can be identified: Dual-porosity models (Fig. 2g), also known as mobile–immobile models (Armstrong et al., 2000), assume that the permeability of the soil matrix is very low due to poorly connected micropores and, thus, water in the matrix is immobile in the vertical direction. In contrast, dual-permeability models (Fig. 2h) allow for slow flux through the matrix in addition to the flux through macropores (Malone et al., 2004; Simunek and van Genuchten, 2008).

For more detailed mathematical descriptions of physically based soil water flow modeling, especially preferential flow, the reader is referred to the reviews of Gerke (2006) and Simunek et al. (2003). Köhne et al. (2009) provide a comprehensive review of applications of different modeling approaches undertaken during the last decade.

Detailed physically based models are primarily applied to studies at the scale of a soil column (length: 0.1 m–1 m) or a plot (soil profile: 1 m<sup>2</sup> to plot: 1000 m<sup>2</sup>) (Köhne et al., 2009). However, if the same approaches are used at hillslope-scale or even catchment-scale studies (ha up to several tens of km<sup>2</sup>), which often use large spatial units (grid cells) to represent the study area, some doubt arises as to whether or not these models can still be referred to as ‘physically based’, as expressed by several authors: “It seems to be an unresolved question if solving overland or groundwater flow equations for horizontal cell sizes of 100 m × 100 m, with average properties, can be considered a physically based approach” (Köhne et al., 2009:16). “The use of the Richards equation at the field and watershed scale is based more on pragmatism than on a sound physical basis” (Vereecken et al., 2007:1). As scale increases, other processes tend to dominate the overall flux or some local-scale effects become subordinate, causing a different behavior of the system. As a result, these systems exhibit complex nonlinear and, sometimes, threshold-type behavior that is difficult to describe with conventional methods (Tetzlaff et al., 2008; Tromp-van Meerveld and McDonnell, 2006).

## 2.2. Soil Property Estimation for Model Parametrization

The major challenges in soil parameter estimation for hydrologic modeling are related to scale issues and spatial variability. In this section, we briefly address (1) the challenge of upscaling soil hydraulic properties based on point- or plot-scale measurements and (2) interpolation or regionalization of point- and plot-scale measurements to locations with unknown soil parameters.

While parameters of soil properties are mainly measured at the scale of a soil core, for modeling estimates of parameter values are needed at larger spatial units such as grid cells or entire catchments. Parameters or models, which are used to simulate the response of large-scale systems without actually accounting for heterogeneity of properties and processes at the small scale are called effective parameters or models (Grayson and Blöschl, 2000). Environmental isotope studies revealed that catchment-scale soil hydraulic properties are different from those at soil core and plot scale (Bazemore et al., 1994). Catchment-average hydraulic conductivity is also found to increase with increasing catchment size (Blöschl and Sivapalan, 1995) and has been observed to be 2–15 times larger on the hillslope scale than values derived from soil core measurements (Brooks et al., 2004). A possible explanation is that with increasing spatial scale, the frequency and connectivity of preferential flow-paths are higher and, thus, average flow velocities are larger than in isolated soil

cores. Therefore, a central question in hydrologic model parametrization is: How can point observations and small-scale process understanding be meaningfully linked to catchment-scale modeling? To obtain this, scaling is needed which is the procedure of determining effective parameters, properties, or mathematical descriptions needed on a target scale by incorporating information from a different scale. Two main categories of upscaling procedures exist (Vereecken et al., 2007): (1) Forward upscaling approaches derive target-scale parameters (in our case the catchment scale) from information on the spatial structure and variability of soil properties measured at a smaller scale (in our case, point or plot scale). (2) Inverse upscaling approaches derive parameters of the large-scale model and its large-scale model domain from (a) large-scale measurements (often remote-sensing data) or (b) from small-scale measurements, which are used as target variables in the calibration process of the model. In the latter case, it is assumed that the effective model structure or model equations describing the large-scale system behavior are known a priori.

Another challenge in model parametrization arises from the fact that we are unable to measure relevant soil parameters at all locations within a catchment over the entire depth of the soil profile. During the last decade, at least attempts have been made to capture the time-variant spatial structure of soil water flow and subsurface runoff processes using dense spatio-temporally distributed measurements: The Tarrawarra soil moisture data set (Australia) (Western and Grayson, 1998) is an example of an extensive measurement campaign to capture spatio-temporal soil moisture conditions of an entire catchment. It comprises more than 10,000 individual soil moisture observations collected during thirteen measuring campaigns within a two-year period. Soil moisture measurements were taken on a spatial resolution of  $10 \times 20$  m within the 10.5-ha catchment using a mobile time-domain-reflectometry (TDR) instrument. Such surveys are expensive and time consuming; therefore, they are rare and only applicable in comparatively small catchments of a few 10th of hectares (Bogena et al., 2010).

Another central question in model parametrization is how to estimate soil parameter values of all locations within a catchment, based on a limited set of measurements. Interpolation and regionalization are tools to perform this task. Interpolation techniques, such as kriging or inverse distance weighting (IDW), use spatial distance as a weighting factor to estimate a parameter value based on neighboring measurement points. In addition, some other techniques incorporate auxiliary variables such as topographic indices or land use to estimate spatial patterns from point measurements (Lyon et al., 2010). Regionalization refers to methods which estimate variable values based on relations between the variable in question and known properties of the point or area in question. The relations have to be established based on observations of variables and properties of other sites prior to the regionalization procedure. In soil science pedotransfer functions are common to estimate soil hydraulic properties based on similarity of mapped soil types or other soil properties, which are easier or

more common to be observed (Lin et al., 2006; Pachepsky et al., 2006). However, the spatially transferred values of soil hydraulic properties are not necessarily representative for catchment-scale effective values as they still might be derived from small-scale core samples (Lin et al., 1999). From a modeler's point of view, this inconsistency between point observations and effective parameters makes it difficult to evaluate whether or not the model is simulating internal runoff processes reasonably well. If careful evaluation and reasoning of the model structure are not undertaken, soil parameters of simple conceptual runoff models might also compensate for structural model errors and correlate with other characteristics in a catchment than those related to soils (Seibert, 1999).

### 3. TACIT KNOWLEDGE OF SUBSURFACE RUNOFF PROCESSES – THE EXPERIMENTALISTS' POINT OF VIEW

Tacit knowledge is expert knowledge that is internalized intuitively through repeated experience or observation (Hudson, 1992). It is inherent to its nature that tacit knowledge is difficult to express explicitly and to transfer from one expert to other persons (Polanyi, 1966). In fact, a major challenge of the dialog between the experimentalists and modelers in hydrologic science is the difficulty of (tacit) knowledge transfer and linking the experimentalists' observations to the governing mechanisms of the modelers' simulations. In the following part of this paper, we discuss a variety of hydric soil indicators, the experimentalists' tacit knowledge, and how to link it to subsurface runoff processes.

#### 3.1. Indicating Dominant Subsurface Runoff Processes Based on Soil Survey Information

In this section, the potential of soil survey information to indicate dominant subsurface runoff processes is going to be discussed. The term “*dominant subsurface runoff processes*” is used for several reasons: (1) Soil survey information is not representative of seldom, temporary processes but is a fingerprint of *dominant*, prevailing, long-term, and persistent soil developing processes. (2) The dominant process concept (Grayson and Blöschl, 2000), which frames the following sections of this paper, was motivated by the recognition of being unable to observe and model all processes in detail, by the knowledge that only a few processes might be dominating the system behavior, and by the experience that simple models with only a few dominating parameters can be successful in modeling catchment runoff response (Sivakumar, 2004). The perceptual model, which comprises of the imagination of how the field evidences fit into dominant processes, is strongly influenced by the experimentalists' tacit knowledge. We begin this section with relevant information on subsurface runoff processes which can be derived from soil

classes and their taxonomy used in soil mapping, continue with properties identifiable at a soil profile, and conclude with a discussion of small-scale hydromorphic features found in individual soil horizons, from which dominant subsurface runoff processes can be inferred. Having described the modelers' point of view on soil hydrologic function in Section 2, this section describes how experimentalists see soils and subsurface runoff processes.

Soil classes and their taxonomy shown in soil maps of various scales allow to indicate the degree and duration as well as the nature of soil saturation within a soil profile: Soils which are classified as gley soils are characterized by soil saturation in all layers from the upper boundary of saturation to a depth of 2 m or more by definition (Soil Survey Staff, 2006). They differ from soils classified as pseudo-gley insofar as they show saturation of one or more layers overlying a water-restricting, unsaturated horizon within 2 m of the surface (Soil Survey Staff, 2006). From this definition it is apparent that in pseudo-gley perched water tables must form and lateral subsurface flow can be expected in the soil horizons above the water-restricting layer (see Fig. 1f). In gley soils subsurface flow can be assumed to occur over the whole depth of the saturated soil profile. Both soil types are likely to show saturation-excess overland flow (Fig. 2b) as the groundwater table is generally high and therefore available storage capacity is rather small. This example shows that information on dominant subsurface runoff processes is inherent in soil classes and spatial distribution of dominant subsurface runoff processes can be roughly derived from soil maps, which might cover large areas of investigation (see also HOST classification in Section 3.2.2). Detailed soil surveying information often covers hydric properties. Natural drainage classes, for instance, describe the frequency and duration of wet periods within a year which have dominated soil formation at a particular site (Soil Survey Division Staff, 1993). Water-state annual patterns are describing soil water states of individual soil horizons or a standard depth zone over a year in tabular form and can show which soil horizons are saturated and therefore conductive at which time of the year. Inundation classes or inundation occurrences describe the degree, frequency, and duration of free water above the soil (Soil Survey Division Staff, 1993). Surface sealing, shrinking cracks, and indicators of hydrophobicity such as waxy coating of soil aggregates after a wild fire are recorded in a soil survey and can be used to anticipate infiltration capacity of soils and potential occurrence of infiltration-excess overland flow (Fig. 1a).

Soil horizons are the result of persistent subsurface flow and associated leaching, transport, and accumulation of material in different soil horizons. From soil hydraulic properties of individual soil horizons such as the saturated hydraulic conductivity – in both lateral and vertical directions – drainable porosity or bulk density, potential dominant subsurface runoff can be inferred. To estimate the storage capacity and the nature of onset of subsurface runoff, the depth to a water-restricting layer or a hydraulic discontinuity as well as

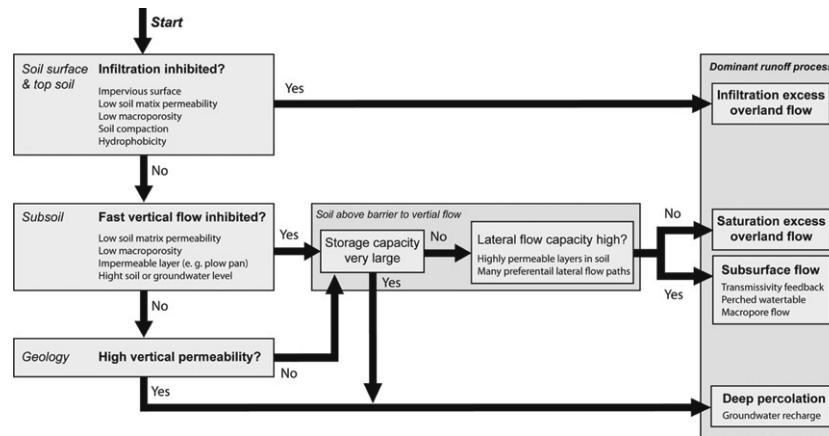
depth to the soil–bedrock interface are important to consider. All this information is included in a description of a soil profile. If soil layers get saturated, highly conductive soil layers or preferential flow paths can be activated and, if connected efficiently, this can lead to a rapid contribution to stream flow discharge in a nearby channel (Seibert et al., 2009; Whipkey, 1965). Saturation occurs most likely at hydraulic discontinuities, either at water-restricting soil horizons (Fig. 1f) or at the soil–bedrock interface (Fig. 1g) (Freer et al., 2002; McDonnell, 1990). Water-restricting soil layers often form due to elluvial and illuvial processes which can be the result of persistent soil water flow driven by the hydraulic gradient and thus driven by topography. Confining flow trajectories due to topography, for instance, are the reason for constant soil water supply in depressions and thus influence soil formation there. For that reason, there is a feedback between topography, soil moisture state, and soil formation, which can be used to anticipate soil attributes at different topographic units within a landscape. A soil catena, for instance, maps this spatial dependency of soil properties and topography along a hillslope. While nowadays information on surface topography is available in detail for almost all catchments in form of digital elevation models (DEMs) there is hardly any information on the subsurface and bedrock topography, which is similarly important for anticipating dominant subsurface runoff processes (Freer et al., 2002).

When looking at individual soil horizons in more detail, a number of small-scale, visible hydric soil indicators can be observed which are characteristic to certain patterns of soil water movement. Soil redox features, also called soil mottling, are an example of soil morphological indicators which form by alternating reduction and oxidation due to saturation and drying which is associated with precipitation of Fe and Mn compounds in the soil (Soil Survey Staff, 1999). Redox concentration around macropores and redox depletion in the soil matrix could suggest that the macropores get drained and aired from time to time while the soil matrix stays more or less saturated constantly (Lin et al., 2008). The Soil Survey Staff at the US Department of Agriculture – National Resources Conservation Service (2010) published a set of such soil morphological characteristics (“field indicators of hydric soils”) which allow a link to long-term persistent flow and transport processes.

### 3.2. Decision Schemes to Identify Dominant Runoff Processes

Linking hydric soil indicators, which can be identified in the course of a soil survey, with dominant runoff processes is often a matter of the experimentalists’ tacit knowledge and thus subjective. Therefore, there is a need for rule-based decision schemes to link soil and land-use properties to dominant subsurface runoff processes in a structured and more objective way (Fig. 3). The spatial distribution of these dominant runoff processes can be expressed in the form of dominant runoff process maps. These maps structure the total area of a catchment into smaller, homogeneous units characterized by one dominant





**FIGURE 3** Schematic example of a decision scheme inferring dominant runoff processes from properties of top soil, subsoil, and geology. The decision tree is a means to capture the experimentalists' tacit knowledge so that it can be applied by others in an objective and thus comparable way. Modified based on Schmocker-Fackel, 2004.

runoff process. In the following, an overview of existing decision schemes structured into bottom-up and top-down approaches (as suggested by Schmocker-Fackel et al., 2007) will be given.

### 3.2.1. Bottom-up Approaches

Bottom-up approaches (e.g. Peschke et al., 1999; Scherrer and Naef, 2003; Schüller, 2005) are based on detailed soil surveying and/or artificial rainfall simulations performed on selected plots in the field. Processes, which have been identified to be dominant in the course of these detailed field investigations, are then assigned to other areas in the catchment having similar properties. To determine similarity, maps regarding soil type, land use, topography, and the drainage network are utilized (Peschke et al., 1999; Schmocker-Fackel et al., 2007). Markart et al. (2004b) present guidelines based on artificial rainfall simulations to derive infiltration capacity or surface runoff coefficients for various types of montane and alpine vegetation, including forest and pasture. The approach is different from the following as it is not considering subsurface processes but only infiltration capacity. However, it is an impressive example of experimentalists' aim to systematically build up process understanding in the form of an exceptional data set of 700 artificial rainfall simulations and to distill it down to a set of rules to indicate infiltration capacity (Markart et al., 2004a). Several decision schemes go further by inferring not only infiltration capacity, but also the dominant runoff processes (including subsurface runoff) from soil and land-cover or land-use properties.

Scherrer and Naef (2003) presented a decision scheme based on plot-scale artificial rainfall simulations and detailed observations of resulting surface and

subsurface runoff processes. The decision scheme is structured like a soil profile considering the soil properties of the surface, topsoil, subsoil, and underlying geology and results in dominant runoff process maps (Fig. 3). This decision scheme was primarily developed for grassland areas and events characterized by high rainfall intensities but was further extended to agricultural land and forested areas incorporating both high and moderate rainfall intensities (Scherrer, 2006). Schmocker-Fackel (2004) and Schmocker-Fackel et al. (2007) simplified the decision scheme of Scherrer and Naef (2003) and proposed a workflow based on geographical information system (GIS) for deriving dominant runoff process maps utilizing existing soil data. Naef et al. (2007) demonstrated that this type of automated scheme allows application beyond the local scale and enables the generation of dominant runoff process maps for areas larger than 1000 km<sup>2</sup>. As these authors admit, detailed geo-information such as soil maps – in their case on a scale of 1:5000 – are not widespread. Therefore, this approach is only applicable in certain data-rich regions. Schüller (2005) developed a similar decision scheme for forested areas using detailed forest management maps available in Rheinland-Pfalz, Germany (1:5000 to 1:10,000 scale). The decision scheme to assign a dominant runoff process is based on Scherrer and Naef (2003) and Scherrer (2006). Schüller (2005) further developed the approach by considering the delay of runoff process response as a function of the slope gradient. Steeper areas are expected to have fast responding surface and subsurface runoff processes, whereas flatter ones are expected to show a delayed response. In addition, line features such as drainage ditches, which can act as efficient connections between areas of runoff formation and the catchment outlet, are taken into account. The decision scheme by Waldenmeyer (2003) is based on detailed forest management maps and incorporates morphometric indices. While Scherrer and Naef (2003) and Schüller (2005) consider the runoff response of mapped dominant runoff processes to be invariant regardless of rainfall intensity, Waldenmeyer (2003) incorporates scenarios of various antecedent system conditions and precipitation intensities to account for varying runoff response.

## 3.2.2. Top-down Approaches

The bottom-up approaches discussed above require detailed soil- and land-use information to regionalize point observations to the catchment scale. Therefore, several authors (e.g. Tilch et al., 2006; Uhlenbrook, 2003) have followed a top-down approach. This type of approach starts with identifying hydrologically homogeneous units (HHU) based on existing geo-data (e.g. coarse DEMs and regional-scale soil and geological maps) or maps derived from aerial photos or other remote-sensing information (e.g. vegetation- and land-use maps). Compared to the maps used in bottom-up approaches, these data sets are generally coarser. HHUs derived by intersecting these data sets with a geographical information system (GIS) are then assumed to have hydrologically similar properties. Uhlenbrook (2003) and Tilch et al. (2006) used

a top-down approach to delineate homogeneous landscape units in mesoscale catchments in Germany and Austria, respectively. Both authors performed field experiments in their catchments to assign hydrologically relevant properties to specific HHUs. Tilch et al. (2006) developed a regionalization approach to estimate additional necessary hydrologically relevant properties of the underlying sediments in the catchment based on DEM analysis. Transferability of such a top-down approach was tested successfully in a neighboring catchment (Uhlenbrook, 2003).

In Great Britain, a top-down approach was developed even applicable to indicate dominant runoff processes of very large areas up to hundred thousands of square kilometers. The Hydrology of Soil Types (HOST) classification (Boorman et al., 1995) is a soil-classification scheme based on regional soil maps of Great Britain (1:250,000 scale), to which different conceptual models of dominant subsurface runoff processes and pathways through the soil profile and/or parent material are assigned. The decision scheme first classifies soils into their physical setting by distinguishing between soils on a permeable parent material with mainly (1) deep groundwater tables, (2) shallow groundwater tables, and (3) soils and parent material with water-restricting layers within 1 m of the surface (Boorman et al., 1995). Within these three settings, 11 HOST response models are defined to account for differences in soil properties and wetness regimes. Each HOST response model can have further subdivisions regarding flow rate or flow mechanism, water-storage capacity, geology of parent material, saturated hydraulic conductivity, or artificial drainage. As not all combinations of HOST response models and subdivisions are realistic to occur and different combinations result in a similar hydrologic behavior, the set is limited to 29 combinations of so-called HOST classes. The HOST decision scheme was optimized by regressing base flow index (BFI) and the standard percentage runoff (SPR) – both characterizing catchment runoff response – to fractions of the various HOST classes occurring within the watershed of selected test catchments (Boorman et al., 1995). This HOST classification was then evaluated performing the same regression based on a set of 575 catchments in Great Britain, which resulted in a coefficient of determination of 0.79 and a standard error of 0.089 (Boorman et al., 1995). The results give reasonable confidence that the HOST classification can be a useful tool to estimate catchment-scale runoff response (BFI and SPR) for a very large area of investigation with lots of catchments incorporating the experts' process understanding. HOST maps with a spatial resolution of  $1 \times 1$  km are available in digital form for the entire United Kingdom (data can be obtained from *Macaulay Land Use Research Institute*). In addition, lookup tables exist to convert more detailed soil maps to HOST classes (Boorman et al., 1995). Schneider et al. (2007) present a HOST-based hydrologic classification for all of Europe based on the Soil Geographical Database of Europe (SGDBE) scaled 1:1 million.

A criticism of the original work of Boorman et al. (1995) is that the HOST-based soil information cannot be directly linked to model parameters that are

commonly used in rainfall–runoff modeling. For that reason, Dunn and Lilly (2001) present an approach linking soil parameters of a conceptual model to HOST classes. Monte Carlo simulations were performed to optimize parameter sets for two Scottish catchments of different character using the same meteorological input. Differences in optimal parameter sets of the two catchments were then assumed to reflect differences in fraction of HOST classes. Therefore, different parameter sets could be assigned to HOST classes, dominating in either the one or the other catchment.

### 3.2.3. *Critical Thoughts and Remaining Challenges*

Having mentioned bottom–up and top–down decision schemes to indicate dominant runoff processes for small-/mesoscale up to large-scale applications we now briefly discuss three critical challenges related to these approaches:

- 1) The results of field experiments that are representative of the governing processes at the point or plot scale are sometimes weak in representing the general hydrologic behavior of a larger scale (e.g. entire catchment). Scherrer and Naef (2003) report different hydrologic responses on soils with rather similar characteristics when performing their artificial rainfall simulations on selected plots. In fact, the influence of key processes emerged only in the context of the actual landscape position when considering the neighboring HHUs.
- 2) For many catchments of interest soil information does not exist and inferring dominant runoff processes from mapped soil indicators would be too cost- and labor-intensive. Therefore, various authors attempted to use indicator variables such as topography or vegetation in a preliminary working step to generate a first draft of a dominant runoff process map, which could then be evaluated in the field in a relatively short period of time (Meissl et al., 2009; Scherrer, 2006; Schmidt et al., 2000). Meissl et al. (2011) showed for an alpine study catchment that for 85% of a validation set of grid cells the same dominant runoff process could automatically be assigned as mapped in the field when using a classification and regression tree approach and the vegetation as the only explanatory variable. When incorporating both vegetation and topography, 98% of the validation set of grid cells could be classified correctly. The digital elevation model (DEM) can also be used to derive the degree of convergence and the slope gradient to estimate soil depths (Dahlke et al., 2009; Hjerdt et al., 2004) and thus storage capacity, which are good indications of potential dominant runoff processes (Scherrer, 2006).
- 3) A systematic comparison of the existing approaches to derive dominant runoff processes has not yet been published. This lack of comparative studies is partly due to the fact that the schemes require different sort and level of detail of geo-input data (soil- and land-use maps, DEM, etc.). Not all schemes differentiate between the same runoff processes and some do not consider

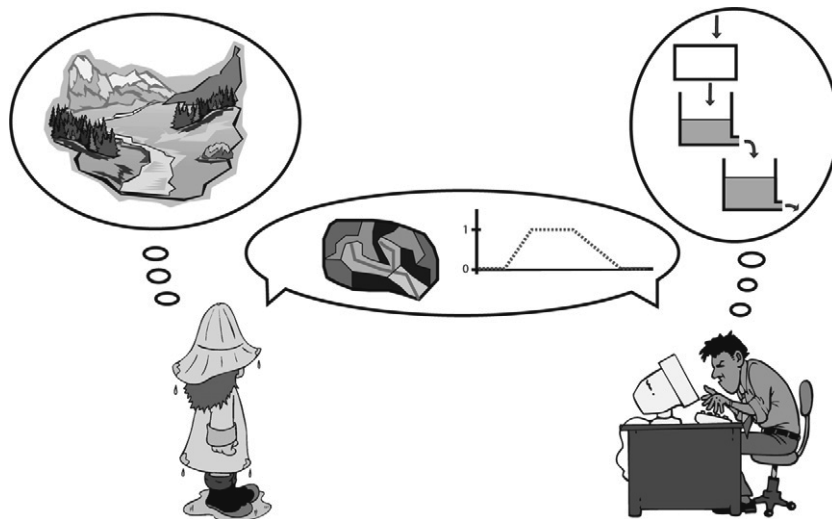
process intensities or runoff delay, respectively. Furthermore, the approaches are developed for different environments and are based on a limited number of field experiments representative for these environmental settings.

## 4. DISCUSSION

Having described the different points of view of modelers and experimentalists on subsurface runoff processes in the previous sections the question remains, how the dialog between experimentalists and modelers can actually be enhanced. In the following, we describe the potential of the Soft Data Concept (Fig. 4) and flexible box models to overcome the difficulty. We start with a general characterization of soft data.

### 4.1. The Soft Data Concept

Quantitative data, also called hard data, in soil surveying such as porosity or saturated hydraulic conductivity can directly be incorporated into quantitative models. In contrast, soft data are more difficult to consider in numerical modeling because of the following features: (1) Soft data are of qualitative, semi-quantitative, or categorical nature (e.g. natural drainage classes, inundation occurrence and hydric soil indicators). Soft data might (2) not meet



**FIGURE 4** The Soft Data Concept as a tool to enhance the dialog between the experimentalists and modelers: Mapping is a means for synthesizing the experimentalists' tacit knowledge, gained in the course of field experience, which the modelers can distill down to an appropriate model structure. As qualitative or semi-quantitative data – called soft data – cannot directly be incorporated in the model calibration/validation procedure, fuzzy-member functions are used to assess the degree of acceptability of modeling results.

official quality standards and therefore cannot directly be utilized to calibrate a model without prior critical assessment: Information on water levels of historic flood events, for instance, which are based on old literature in archives (Schmocker-Fackel and Naef, 2010) or which are affected by unreliable rating curves, can still be very informative about the potential flood magnitude of extreme events. These rare events with a low probability of occurrence are normally not covered in existing time series of modern stream flow discharge measurements. Soft data might (3) first need the experts' interpretation to be applicable to modeling: Tracer observations, for instance, allow estimating catchment-scale mean transit times, which can be valuable for characterizing the hydrologic functioning of a catchment (Soulsby et al., 2010). Soft data can also (4) be affected by considerable spatial variability, often not resolved by the coarse spatial discretization of hydrologic models (subgrid variability): Data from point measurements such as groundwater levels or soil moisture data, used to calibrate and validate a model or its internal model state (Freer et al., 2004), are known to be affected by small-scale topographic and pedologic properties and therefore might not be directly comparable with modeling results on a coarser spatial resolution. The time series at hand might (5) not be continuous but only cover some events measured during a few field campaigns, thus the data are called "soft". Soft data might (6) also include indirect measurements or derived values of catchment-scale system states. Such soft information could be based on remotely-sensed data or upscaled from direct measurements using spatial interpolation, regionalization or other computational methods. Thermal images, for instance, might be useful to identify spatio-temporal patterns of soil moisture distribution or evaporation within a catchment (Ludwig et al., 2003). Another example is an estimate of catchment-wide storage based on field measurements, which can inform subsurface models (Seibert et al., 2011). Finally, it is important to recognize that (7) the experts' tacit knowledge is qualitative and subjective and therefore much harder to directly link to quantitative model results in an objective way.

Despite these limitations, soft data are particularly valuable for constraining model parameters and optimizing model structure in the calibration/validation process. In addition to quantitative goodness-of-fit criteria like the Nash–Sutcliffe efficiency (Nash and Sutcliffe, 1970), soft data can be useful to define a realistic range of possible parameter values in an automated calibration procedure (Winsemius et al., 2009). Soft data can be used to assess the acceptability of model simulations or parameter values. When the degree of acceptance is high the parameter values agree well with the field experience of the experimentalists. Fuzzy-logic based rules are applied to transform the qualitative information into a numeric value of acceptance (Seibert and McDonnell, 2002). For each type of soft data, which is going to be considered in the model calibration/validation procedure, a fuzzy-member function (Fig. 4) is defined to calculate the degree of acceptance (ranging from 0 to 1) of model results based on the experimentalists' tacit knowledge how the system should

function. When using multiple soft data sources at once, the total degree of acceptance can be determined by calculating a mean (e.g. geometric mean) of individual acceptability scores. The goal of the model calibration procedure is then not only to result in a high goodness-of-fit score of modeled total catchment runoff but also to optimize the degree of acceptability of the internal model processes and, thus, to better agree with what the experimentalists observe in the field. Flexible box models, a type of conceptual model with tunable storage properties and components, allow optimizing the model structure to best capture all necessary processes identified by the experimentalists (Fenicia et al., 2008; Schmocker-Fackel, 2004; Uhlenbrook and Leibundgut, 2002). Soft data become particularly valuable for model exercises in ungauged catchments, where no hard data exist. In short, there may be a wealth of untapped data available for many catchments, which have not been utilized until now.

## 5. CONCLUDING REMARKS

For future progress in hydropedology a dialog between experimentalists and modelers offers great potential as both provide different views on investigating, understanding, and quantifying the first-order control of catchment runoff response. The experimentalists can offer a wealth of hard and soft facts, documented in soil surveys as well as their tacit knowledge gained during many years of field experience. A set of hydric soil indicators and decision schemes exist to systematically indicate dominant subsurface runoff processes. Mapping these dominant runoff processes has been identified as a potential way to capture the experimentalists' tacit knowledge and transfer it to the modelers, who are then able to give feedback based on their modeling results. The modelers can inform model calibration and parameter estimation in the course of the Soft Data Framework. This allows incorporating qualitative and semi-quantitative data, existing for not only gauged but also ungauged catchments. In this way, acceptability of model simulations is evaluated not only based on catchment runoff as a single integrated value but also based on additional information about internal system states and processes. The dialog between experimentalists and modelers should not necessarily aim at highly detailed observations and modeling of natural heterogeneity with evermore complex monitoring and modeling techniques. Instead, the goal should be to reveal first-order controls of natural hydrologic systems to better understand nonlinearity of runoff processes at the catchment scale (McDonnell, 2003).

## ACKNOWLEDGMENTS

The authors thank Henry Lin, the editor of this book, Gertraud Meissl, Russell Smith, and two anonymous reviewers for their critical reviews and useful comments as well as Christine Rinderer for proofreading the manuscript.

## REFERENCES

- Armstrong, A.C., Matthews, A.M., Portwood, A.M., Leeds-Harrison, P.B., Jarvis, N.J., 2000. CRACK-NP: a pesticide leaching model for cracking clay soils. *Agr. Water Manage.* 44 (1–3), 183–199.
- Bazemore, D.E., Eshleman, K.N., Hollenbeck, K.J., 1994. The role of soil water in stormflow generation in a forested headwater catchment: synthesis of natural tracer and hydrometric evidence. *J. Hydrol.* 162 (1–2), 47–75.
- Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modeling - a review. *Hydrol. Process.* 9 (3–4), 251–290.
- Bogena, H.R., Herbst, M., Huisman, J.A., Rosenbaum, U., Weuthen, A., Vereecken, H., 2010. Potential of wireless sensor networks for measuring soil water content variability. *Vadose Zone Journal* 9 (4), 1002. Doi:10.2136/vzj2009.0173.
- Boorman, D., Hollis, J., Lilly, A., 1995. Hydrology of Soil Types: a Hydrologically Based Classification of the Soils of the United Kingdom. Report 126. Institute of Hydrology, Wallingford.
- Brooks, E.S., Boll, J., McDaniel, P.A., 2004. A hillslope-scale experiment to measure lateral saturated hydraulic conductivity. *Water Resour. Res.* 40 (4). Doi: 10.1029/2003wr002858.
- Dahlke, H.E., Behrens, T., Seibert, J., Andersson, L., 2009. Test of statistical means for the extrapolation of soil depth point information using overlays of spatial environmental data and bootstrapping techniques. *Hydrol. Process.* 23 (21), 3017–3029. Doi: 10.1002/Hyp.7413.
- Dunn, S.M., Lilly, A., 2001. Investigating the relationship between a soils classification and the spatial parameters of a conceptual catchment-scale hydrological model. *J. Hydrol.* 252 (1–4), 157–173.
- Fenicia, F., McDonnell, J.J., Savenije, H.H.G., 2008. Learning from model improvement: on the contribution of complementary data to process understanding. *Water Resour. Res.* 44 (6). Doi: 10.1029/2007wr006386.
- Freer, J., McMillan, H., McDonnell, J.J., Beven, K.J., 2004. Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. *J. Hydrol.* 291 (3–4), 254–277.
- Freer, J., McDonnell, J.J., Beven, K.J., Peters, N.E., Burns, D.A., Hooper, R.P., Aulenbach, B., Kendall, C., 2002. The role of bedrock topography on subsurface storm flow. *Water Resour. Res.* 38 (12). Doi: 10.1029/2001wr000872.
- Gerke, H.H., 2006. Preferential flow descriptions for structured soils. *J. Plant Nutr. Soil Sci.* 169 (3), 382–400. Doi: 10.1002/jpln.200521955.
- Grayson, R., Blöschl, G., 2000. Spatial Patterns in Catchment Hydrology: Observations and Modelling. Cambridge University Press, Cambridge.
- Hjerdt, K.N., McDonnell, J.J., Seibert, J., Rodhe, A., 2004. A new topographic index to quantify downslope controls on local drainage. *Water Resour. Res.* 40 (5). Doi: 10.1029/2004WR003130.
- Hua, J.P., Liang, Z.M., Yu, Z.B., 2003. A modified rational formula for flood design in small basins. *J. Am. Water Resour. Assoc.* 39 (5), 1017–1025.
- Hudson, B.D., 1992. The soil survey as paradigm-based science. *Soil Sci. Soc. Am. J.* 56 (3), 836–841.
- Kirchner, J.W., 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.* 42 (3). Doi: 10.1029/2005wr004362.
- Klemes, V., 1986. Dilettantism in hydrology - transition or destiny. *Water Resour. Res.* 22 (9), 177–188.



- Köhne, J.M., Köhne, S., Simunek, J., 2009. A review of model applications for structured soils: (a) water flow and tracer transport. *J. Contam. Hydrol.* 104 (1–4), 4–35. Doi: 10.1016/j.jconhyd.2008.10.002.
- Lin, H., Brooks, E., McDaniel, P., Boll, J., 2008. Hydropedology and surface/subsurface runoff processes. In: Anderson, M.G. (Ed.), *Encyclopedia of Hydrological Sciences*. Wiley, Chichester, Online Library, pp. 1–25.
- Lin, H., Bouma, J., Pachepsky, Y., Western, A., Thompson, J., van Genuchten, R., Vogel, H.J., Lilly, A., 2006. Hydropedology: synergistic integration of pedology and hydrology. *Water Resour. Res.* 42 (5). Doi: 10.1029/2005wr004085.
- Lin, H., McInnes, K.J., Wilding, L.P., Hallmark, C.T., 1999. Effects of soil morphology on hydraulic properties: II. Hydraulic pedotransfer functions. *Soil Sci. Soc. Am. J.* 63 (4), 955–961.
- Ludwig, R., Probeck, M., Mauser, W., 2003. Mesoscale water balance modelling in the Upper Danube watershed using sub-scale land cover information derived from NOAA-AVHRR imagery and GIS-techniques. *Physics and Chemistry of the Earth* 28 (33–36), 1351–1364.
- Lyon, S.W., Sorensen, R., Stendahl, J., Seibert, J., 2010. Using landscape characteristics to define an adjusted distance metric for improving kriging interpolations. *Int. J. Geogr. Inf. Sci.* 24 (5), 723–740. Doi: 10.1080/13658810903062487.
- Malone, R.W., Ahuja, L.R., Ma, L.W., Wauchope, R.D., Ma, Q.L., Rojas, K.W., 2004. Application of the root zone water quality model (RZWQM) to pesticide fate and transport: an overview. *Pest Manag. Sci.* 60 (3), 205–221. Doi: 10.1002/Ps.789.
- Markart, G., Kohl, B., Sotier, B., Schauer, T., Bunza, G., Stern, R., 2004b. Provisorische Geländeanleitung zur Abschätzung des Oberflächenabflussbeiwertes auf Alpenen Boden-/Vegetationseinheiten bei Konvektiven Starkregen (Version 1.0) – A Simple Code of Practice for Assessment of Surface Runoff Coefficients for Alpine Soil-/Vegetation Units in Torrential Rain (Version 1.0). BFW-Dokumentationen 3/2004. Bundesamt und Forschungszentrum für Wald, Vienna.
- Markart, G., Kohl, B., Schauer, T., Sotier, B., Bunza, G., Stern, R., 2004a. Eine einfache Gelaendeanleitung zur Abschätzung des Oberflächenabflussbeiwertes bei Starkregen. A simple code of practice for assessment of surface runoff coefficients in torrential rain. pp. 89–100. In: *International Symposium INTERPRAEVENT* (Ed.), *International Symposium INTERPRAEVENT*. 24.05. – 28.05. 2004. Riva del Garda/Trient/Italy.
- McDonnell, J.J., 1990. A rationale for old water discharge through macropores in a steep, humid catchment. *Water Resour. Res.* 26 (11), 2821–2832.
- McDonnell, J.J., 2003. Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response. *Hydrol. Process.* 17 (9), 1869–1875.
- McGlynn, B.L., McDonnell, J.J., Brammer, D.D., 2002. A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, New Zealand. *J. Hydrol.* 257 (1–4), 1–26.
- Meissl, G., Geitner, C., Tusch, M., Stötter, J., 2009. Derivation of Dominant Runoff Processes in an Alpine Catchment using Vegetation and Relief Data. *EGU General Assembly 2009*. 19.04 – 24.04. 2009. Vienna.
- Meissl, G., Geitner, C., Tusch, M., Schoeberl, F., Stoetter, J., 2011. Reliefparameter und abflusssteuernde flaecheigenschaften: statistische analyse ihres zusammenhangs in einem kleinen alpinen einzugsgebiet. *Zeitschrift für Geomorphologie* 55 (3), 295–315.
- Mulvaney, T.J., 1851. On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharge in a given cathement. *P. I. Civil Eng. Ireland* (4), 19–31.

- Naef, F., Margreth, M., Schmocker-Fackel, P., Scherrer, S., 2007. Automatisch hergeleitete abflussprozesskarten - ein neues werkzeug zur abschätzung von hochwasserabflüssen. *Wasser Energie Luft* 99 (3), 267–272.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I – a discussion of principles. *J. Hydrol.* 10 (3), 282–290.
- Pachepsky, Y.A., Rawls, W.J., Lin, H.S., 2006. Hydropedology and pedotransfer functions. *Geoderma* 131 (3–4), 308–316. Doi: 10.1016/j.geoderma.2005.03.012.
- Peschke, G., Etzenberg, C., Müller, G., Töpfer, J., Zimmermann, S., 1999. Das Wissensbasierte System FLAB – Ein Instrument zur Rechnergestützten Bestimmung von Landschaftseinheiten mit Gleicher Abflussbildung. *IHI-Schriften* 10. Internationales Hochschulinstitut Zittau, Zittau.
- Polanyi, M., 1966. *The Tacit Dimension*. Doubleday and Co., New York.
- Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. *Journal of Applied Physics* 1 (5), 318–333.
- Rinderer, M., Jenewein, S., Senfter, S., Rickenmann, D., Schöberl, F., Stötter, J., Hegg, C., 2009. Runoff and bedload transport modelling for flood hazard assessment in small alpine catchments-the PROMABGIS model. In: Veulliet, E., Stötter, J., Weck-Hannemann, H. (Eds.), *Sustainable Natural Hazard Management in Alpine Environments*. Springer, Berlin, pp. 69–101.
- Scherrer, S., 2006. Bestimmungsschlüssel zur Identifikation von Hochwasserrelevanten Flächen. Report 18/2006. Landesamt für Umwelt, Wasserwirtschaft und Gewerbeaufsicht, Rheinland-Pfalz.
- Scherrer, S., Naef, F., 2003. A decision scheme to indicate dominant hydrological flow processes on temperate grassland. *Hydrol. Process.* 17 (2), 391–401. Doi: 10.1002/Hyp.1131.
- Schmidt, J., Hennrich, K., Dikau, R., 2000. Scales and similarities in runoff processes with respect to geomorphometry. *Hydrol. Process.* 14 (11–12), 1963–1979.
- Schmocker-Fackel, P., 2004. A Method to Delineate Runoff Processes in a Catchment and Its Implications for Runoff Simulations. PhD-thesis. Swiss Federal Institute of Technology ETH Zurich, Zurich.
- Schmocker-Fackel, P., Naef, F., 2010. Changes in flood frequencies in Switzerland since 1500. *Hydrol. Earth Syst. Sci. Discuss.* 7 (1), 529–560.
- Schmocker-Fackel, P., Naef, F., Scherrer, S., 2007. Identifying runoff processes on the plot and catchment scale. *Hydrol. Earth Syst. Sci.* 11 (2), 891–906.
- Schneider, M., Brunner, F., Hollis, J., Stamm, C., 2007. Towards a hydrological classification of European soils: preliminary test of its predictive power for the base flow index using river discharge data. *Hydrol. Earth Syst. Sci.* 11 (4), 1501–1513.
- Schüler, G., 2005. Herleitung von abflussrelevanten Flächen zur Steuerung von Wasserrückhaltmassnahmen im Wald. *Freiburger Forstliche Forschung* 62. Freiburg im Breisgau.
- Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. *Agr. Forest Meteorol.* 98 (9), 279–293.
- Seibert, J., Bishop, K., Nyberg, L., Rodhe, A., 2011. Water storage in a till catchment. I: Distributed modelling and relationship to runoff. *Hydrol. Process.* 25, 3937–3949.
- Seibert, J., Rodhe, A., Bishop, K., 2003. Simulating interactions between saturated and unsaturated storage in a conceptual runoff model. *Hydrol. Process.* 17 (2), 379–390. Doi: 10.1002/Hyp.1130.
- Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resour. Res.* 38 (11). Doi: 10.1029/2001wr000978.
- Seibert, J., Grabs, T., Kohler, S., Laudon, H., Winterdahl, M., Bishop, K., 2009. Linking soil- and stream-water chemistry based on a Riparian flow-concentration integration model. *Hydrol. Earth Syst. Sci.* 13 (12), 2287–2297.

- Sherman, L., 1932. Streamflow from rainfall by the unit-graph method. *Eng. News Rec.* 108, 501–505.
- Simunek, J., van Genuchten, M.T., 2008. Modeling nonequilibrium flow and transport processes using HYDRUS. *Vadose Zone J.* 7 (2), 782–797. Doi: 10.2136/Vzj2007.0074.
- Simunek, J., Jarvis, N.J., van Genuchten, M.T., Gardenas, A., 2003. Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone. *J. Hydrol.* 272 (1–4), 14–35. Doi: Pii S0022-1694(02)00252-4.
- Sivakumar, B., 2004. Dominant processes concept in hydrology: moving forward. *Hydrol. Process.* 18 (12), 2349–2353. Doi: 10.1002/Hyp.5606.
- Soil Survey Division Staff, 1993. *Soil Survey Manual*. United States Department of Agriculture – Soil Conservation Service, Washington D.C.
- Soil Survey Staff, 1999. *Soil Taxonomy: a Basic System of Soil Classification for Making and Interpreting Soil Surveys*. United States Department of Agriculture - Natural Resources Conservation Service, Washonton D.C.
- Soil Survey Staff, 2006. *Keys to Soil Taxonomy*. United States Department of Agriculture – Natural Resources Conservation Service, Washington, D.C.
- Soil Survey Staff, 2010. *Field Indicators of Hydric Soils in the United States. A Guide for Identifying and Delineating Hydric Soils*. Version 7.0 ed. United States Department of Agriculture – Natural Resources Conservation Service, Washington D.C.
- Soulsby, C., Tetzlaff, D., Hrachowitz, M., 2010. Are transit times useful process-based tools for flow prediction and classification in ungauged basins in montane regions? *Hydrol. Process.* 24 (12), 1685–1696. Doi: 10.1002/Hyp.7578.
- Tetzlaff, D., McDonnell, J.J., Uhlenbrook, S., McGuire, K.J., Bogaart, P.W., Naef, F., Baird, A.J., Dunn, S.M., Soulsby, C., 2008. Conceptualizing catchment processes: simply too complex? *Hydrol. Process.* 22 (11), 1727–1730. Doi: 10.1002/Hyp.7069.
- Tilch, N., Zillgens, B., Uhlenbrook, S., Leibundgut, C., Kirnbauer, R., Merz, B., 2006. GIS-gestützte Ausweisung von hydrologischen Umsatzräumen und Prozessen im Löhnersbach-Einzugsgebiet (Nördliche Grauwackenzone, Salzburger Land). *Österreichische Wasser-und Abfallwirtschaft* 58 (9), 141–151.
- Tromp-van Meerveld, H.J., McDonnell, J.J., 2006. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resour. Res.* 42 (2). Doi: 10.1029/2004wr003800.
- Uhlenbrook, S., 2003. An empirical approach for delineating spatial units with the same dominating runoff generation processes. *Phys. Chem. Earth* 28 (6–7), 297–303. Doi: 10.1016/S1474-7065(03)00041-X.
- Uhlenbrook, S., Leibundgut, C., 2002. Process-oriented catchment modelling and multiple-response validation. *Hydrol. Process.* 16 (2), 423–440.
- Vereecken, H., Kasteel, R., Vanderborght, J., Harter, T., 2007. Upscaling hydraulic properties and soil water flow processes in heterogeneous soils: a review. *Vadose Zone J.* 6 (1), 1–28. Doi: 10.2136/Vzj2006.0055.
- Waldenmeyer, G., 2003. *Abflussbildung und regionalisierung in einem forstlich genutzten einzugsgebiet (Dürreychtal, Nordschwarzwald)*. *Karlsruher Schriften zur Geographie und Geoökologie* 20. Karlsruhe.
- Weiler, M., McDonnell, J., 2004. Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology. *J. Hydrol.* 285 (1–4), 3–18. Doi: 10.1016/S0022-1694(03)00271-3.
- Weiler, M., McDonnell, J.J., Tromp-Van Meerveld, I., Uchida, T., 2005. Subsurface stormflow. In: Anderson, G.M., McDonnell, J.J. (Eds.), *Encyclopedia of Hydrological Sciences*, pp. 1719–1732. Wiley, Chichester.

# Author's personal copy

536

PART | III Advances in Modeling, Mapping, and Coupling

- Western, A.W., Grayson, R.B., 1998. The tarrawarra data set: soil moisture patterns, soil characteristics, and hydrological flux measurements. *Water Resour. Res.* 34 (10), 2765–2768.
- Whipkey, R.Z., 1965. Subsurface stormflow from forested slopes. *Bulletin of the International Association of Scientific Hydrology* 10 (2), 74–85.
- Winsemius, H., Schaefli, B., Montanari, A., Savenije, H., 2009. On the calibration of hydrological models in ungauged basins: a framework for integrating hard and soft hydrological information. *Water Resour. Res.* 45 (12). Doi: 10.1029/2009WR007706.

## PAPER II



# Sensing with boots and trousers – qualitative field observations of shallow soil moisture patterns

Michael Rinderer,<sup>1\*</sup>  
Andrea Kollegger,<sup>1</sup>  
Benjamin M. C. Fischer,<sup>1</sup>  
Manfred Stähli<sup>2</sup> and  
Jan Seibert<sup>1,3,4</sup>

<sup>1</sup> Department of Geography, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland

<sup>2</sup> Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

<sup>3</sup> Department of Earth Sciences, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden

<sup>4</sup> Department of Physical Geography and Quaternary Geology, Stockholm University, SE-10691 Stockholm, Sweden

\*Correspondence to: Michael Rinderer, Department of Geography, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland.  
E-mail: michael.rinderer@geo.uzh.ch

## Abstract

While soil moisture patterns can be interesting traits to investigate spatio-temporal heterogeneity of catchments relevant for various physical processes of soil–atmosphere interaction and soil water redistribution, many of the existing methods to capture spatial patterns are time consuming, expensive or need site-specific calibration. In this study we present a quick and inexpensive supplementary field method for classifying soil wetness in wet environments. The seven wetness classes are based on qualitative indicators, which one can touch, hear or see on the soil surface. To counter critics that such qualitative methods are considerably affected by subjectivity, we performed systematic testing of the method by taking qualitative measurements in the field with 20 non-expert raters. We then analyzed these in terms of degree of agreement and assessed the results against gravimetric sampling and time domain reflectometry measurements. In 70% of all classifications raters agreed on the wetness class assigned to the marked sampling locations and in 95% they were not off by more than one wetness class. The seven quantitative wetness classes agreed with gravimetric and time domain reflectometry measurements, although intermediate to wet classes showed an overlap of their range whereas the driest classes showed considerable spread. Despite some potential to optimize the method, it has been shown to be a reliable supplement to existing quantitative techniques for assessing soil moisture patterns in wet environments. Copyright © 2012 John Wiley & Sons, Ltd.

**Key Words** soil moisture; qualitative field method; wetness classification

## Introduction

The analysis of spatial patterns of different hydrologically relevant variables is a promising way forward to understanding the hydrological functioning of catchments (Sivapalan, 2005; McDonnell *et al.*, 2007). Spatial patterns of soil moisture have been used to identify local and non-local controls of redistribution of water in the unsaturated soil profile (Grayson *et al.*, 1997; Blume *et al.*, 2009; Western and Grayson, 2000). Soil moisture patterns can be indicators of transient saturation and associated subsurface flow processes in soils with limited storage capacities (Western *et al.*, 2004; Western *et al.*, 2005). In such environments, soil moisture patterns have been identified to be predictors of catchment runoff response when considering hydrological connectivity (Western *et al.*, 2001; McNamara *et al.*, 2005; Ali *et al.*, 2010). Furthermore, maps of soil moisture patterns, including saturated areas, are potentially valuable in model validation and calibration (Beven and Kirkby, 1979; Blazkova *et al.*, 2002; Güntner *et al.*, 2004). Capturing these spatial patterns requires spatially distributed measurements. Recently, several studies have been based on the installation of hundreds of spatially distributed soil moisture sensors (e.g. Boga *et al.*, 2010) or have performed repeated measuring campaigns for hundreds of sampling points (Western and Grayson, 1998; Walker *et al.*, 2001; Williams *et al.*, 2003; Meyles *et al.*, 2003; Petrone *et al.*, 2004; Western *et al.*, 2004). However, these measurement approaches are both expensive and time consuming. Moreover, in soils with high humus content, prevailing conditions close to saturation, high porosity and thus low dry bulk density like peat, common measurement techniques, such as time domain reflectometry (TDR), often reach their

Received 10 January 2012

Accepted 23 August 2012

applicability limit or need special calibration to operate properly (Boelter, 1968; Pumpanen and Ilvesniemi, 2005; Michel, 2010).

In related disciplines, qualitative field methods are well established for practical field application. These methods are based on qualitative criteria, which are easily recognizable (i.e. through sight, sound or touch) without additional measuring devices and allow for some assessment or classification of the quantitative value of interest. In soil science, for instance, the texture of a soil (i.e. its content of sand, silt and clay) is estimated by methods such as forming a ribbon with a diameter of roughly half a centimetre and assessing the maximum length that can be obtained before the ribbon is breaking apart (Thien, 1979). In avalanche risk assessment, a quantitative range of snow layer hardness can be estimated depending on whether the snow layer can be penetrated by a fist, a finger, the tip of a pencil or a blade of a knife (De Quervain, 1950; cited in Pielmeier and Schneebeli, 2003). In ecology, the presence of fauna serves as a guide for water quality because they are sensitive to their habitat conditions (Metcalf-Smith, 1994). In ecohydrology, specific plant species are reliable indicators for long-term and prevailing soil wetness conditions (Ellenberg, 1991).

We found surprisingly little in the literature on the use of qualitative criteria to capture shallow soil moisture differences. The Natural Resources Conservation Service of the United States Department of Agriculture (1998) published a photo guide for estimating soil moisture by feel and appearance for four different soil types. A range of quantitative values of available soil moisture remaining in the sample is given for each photo. However, it remains unclear how these values were determined. The influence of user's experience on the estimated soil moisture content is mentioned but not systematically examined. Blazkova *et al.* (2002) defined a qualitative classification scheme based on five wetness classes. However, they did not use the range of these individual wetness classes, rather they aggregated the three wettest ones to assess soil saturation along several transects. Also, they did not evaluate the potential subjectivity in their mapping approach.

In this manuscript, we present a new wetness classification scheme for field application based on qualitative wetness criteria especially developed for wet environmental conditions. In applied studies and pilot investigations, qualitative methods are particularly appealing as they can

supplement quantitative methods. One of the main reasons qualitative methods have not yet been widely accepted is due to their subjectivity and the difficulty of linking qualitative indicators to quantitative values. To test the potential of a qualitative wetness classification scheme, we address the following questions: (i) Is the qualitative wetness classification scheme capable of reliably capturing differences in shallow soil moisture content in a wet environment? (ii) To what extent is the qualitative wetness classification scheme affected by subjectivity? In particular, (iia) how well do the results of wetness classifications by non-expert raters agree? (iib) Are there differences in agreement between the different wetness classes? (iic) Do some individual raters show a systematic difference to the rest of the rating group? (iii) Are the qualitatively defined wetness classes representative of quantitative differences in soil moisture content?

## Methods

### Classification scheme

The wetness classification scheme presented here is based on seven qualitatively defined classes ranging from the driest class (class 1), which would allow a person to sit down on the soil surface without getting wet, to the wettest class (class 7), which is characterized by surface ponding (Table I). The other classes differentiate grades of wetness between these two extremes: Wetness class 6 is defined by water squeezing out of the top most soil when stepping on it with a boot. If only a squelchy noise can be heard when stepping on the ground but no water exfiltrates, then a site would be classified as class 5. Class 4 would not allow a person to sit on the ground without immediately getting wet trousers. At a location classified as class 3, the trousers of a person sitting on the ground would get wet after some minutes, whereas at class 2 locations, they would only get moist but not wet in the same time span. Obviously, it is not intended that a rater using the classification scheme actually sits down on the surface each time to assess the wetness conditions of a sampling point. Rather, it is assumed that most people have some experience and know, or can imagine, whether they would stay dry or get wet if they were to sit on the ground. As vegetation and litter layer potentially influence the rater's class assignments, they should be

Table I. Definition of the qualitative wetness classes

Class	Qualitative indicator criteria
1	The trousers of a person sitting on the ground would stay dry
2	The trousers of a person sitting on the ground would get moist after some minutes
3	The trousers of a person sitting on the ground would get wet after some minutes
4	The trousers of a person sitting on the ground would get wet immediately
5	Squelchy noise can be heard when stepping on the ground but no water is visible
6	Water squeezes out of the topsoil when stepping on it with a boot
7	Water can be seen on the soil surface



bent aside or be removed (see also Discussion and Concluding Remarks section). Rainfall during the sampling campaign as well as dew in the morning might lead to misleading wetness class assignments and thus taking samples at these times should be avoided.

### Datasets and test layout

The field location chosen for testing the qualitative classification scheme is the Erlenbach catchment (Figure 1) in the Alpthal, a pre-alpine valley approximately 40 km southeast of Zurich (Switzerland). In the catchment, 0.5–2 m deep umbric or mollic gleysols with Muck Humus and Mor Humus topsoil have formed on top of a marly parent material (Schleppi *et al.*, 1998). Because of wet environmental conditions, with 2300 mm precipitation a year (Feyen *et al.*, 1999), moor landscapes can be found with a thick organic layer, rich in humus and dense roots. Non-forested locations are generally classified as wetter sampling points. Drier sampling points were either located on steeper, convex slopes or in areas with a light forest cover (predominantly Norway Spruce).

For a proof of concept, we used three different datasets: To assess the capability of the scheme to capture shallow soil moisture differences and to test its sensitivity to subjective influences, a test with 20 master students (geography, fourth year) was performed on 23 May 2011 (dataset 1). Light rainfall with a sum of 1 mm during night before the test and 18 mm during the preceding 3 days had led to moist surface conditions. May 2011 was characterized by repeated rainfall events (16 days with rainfall), resulting in a total sum of 203 mm [long-term average (1982–2010): 212 mm/month].

Choosing non-expert raters, who had not mapped/measured soil moisture in the field before, should minimize the influence of difference in experience of individuals. The students were split into two groups of ten persons, one group classifying a set of 52 sampling points in the morning (group 1) and the other group classifying the same locations in the afternoon (group 2). The change of shallow soil moisture conditions during the day prohibited the two datasets being combined.

To assess the agreement between the qualitative wetness classes of the scheme and the actual soil water content, a different set of 45 sampling points was selected for gravimetric soil water content analysis (dataset 2). Soil samples were taken for all of these locations with a 100 cm<sup>3</sup> steel cylinder in 10 cm depth, and soil wetness was classified by three persons using the qualitative scheme. To further assess agreement between qualitative wetness indicators and quantitative soil water content, a third dataset was analyzed (dataset 3), which consisted of TDR measurements and qualitative wetness classifications from eight sampling campaigns between August and October 2010 (Figure 2). One hundred marked sampling points were randomly distributed over selected plots within the Erlenbach catchment (Figure 1), with a spacing of approximately 10 m. On each day of the campaign, they were assessed by an experienced person applying first the wetness classification and then taking five measurements with a portable TDR device (TRIME-PICO 64) at each sampling location (Kollegger, 2010). Because the same person performed both measurements, dataset 3 potentially could be influenced by a training effect as TDR measurement values of previous sampling points were known. However, the differences between

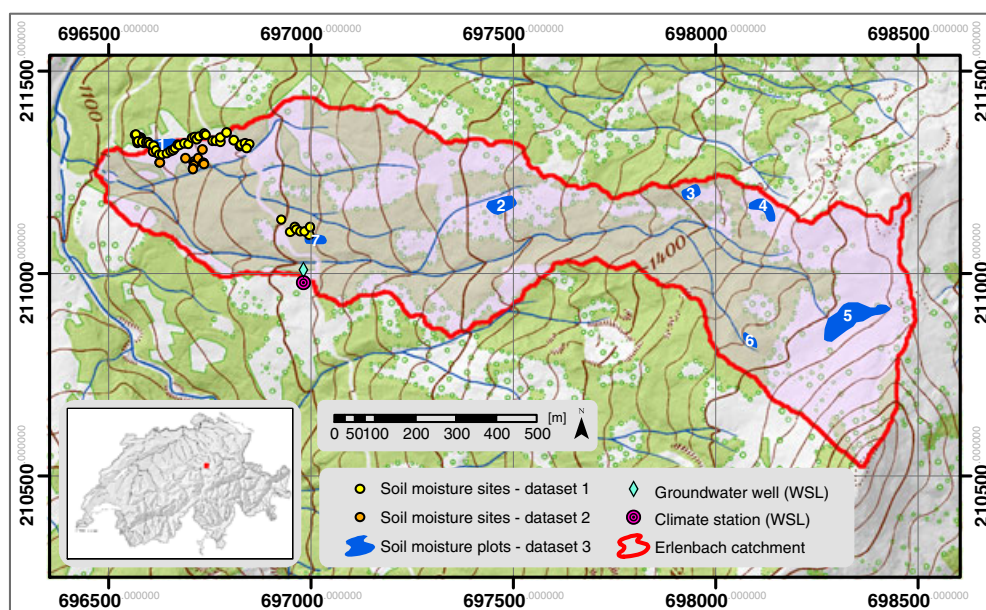


Figure 1. Experimental catchment “Erlenbach” showing sampling locations of all three datasets used in this study as well as the location of the groundwater well and climate station. Note: exact sampling locations of dataset 3 cannot be displayed at the presented scale (topographic map: reproduced under permission of swisstopo (BA12059))

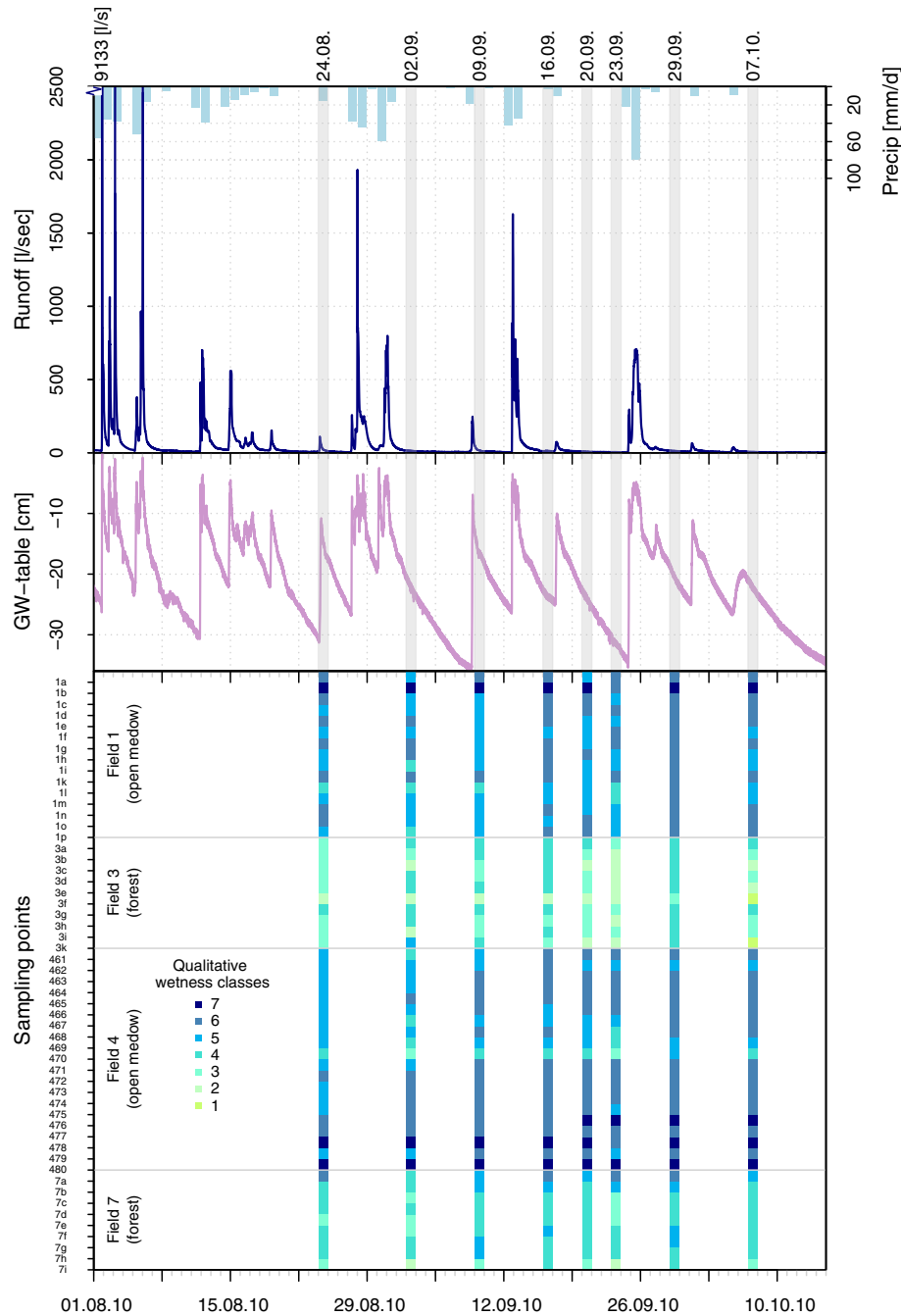


Figure 2. Precipitation, runoff and groundwater levels during summer and fall 2010. Days of moisture sampling campaign are indicated by grey bars in these time series. The coloured rectangles show the observed variations of soil moisture in space and time

volumetric water content measured by TDR and the qualitative classification were approximately constant when comparing the different sampling days as well as for the first and the second half of all sampling points during individual days. This indicated that there was no implicit learning effect. To prevent outliers from influencing the quality of the result, the standard deviation of these five TDR measurements was expected to be smaller than 10% volumetric water content. In cases where this criterion was not fulfilled, the TDR measurements showing the largest deviation from the mean were successively

excluded until the criterion was fulfilled. In cases where less than three measurements were finally available to calculate the mean, the entire sampling point was excluded from the dataset. This selection procedure resulted in 454 data records of TDR and qualitative classification pairs.

## Results

To assess the overall performance of the qualitative classification scheme, the deviation of wetness classification

relative to the mode of each sampling point was analyzed and plotted as frequency distribution (Figure 3). The mode was chosen as reference as it represents the wetness class with the highest frequency of class assignments. In a few cases, two or more classes had exactly the same frequency and the median was then considered to be a more representative reference than arbitrarily choosing one of the two evenly frequent wetness classes. In approximately 70% of all cases classified by raters in each group, they independently assigned the same wetness class. In approximately 95% of all cases, the raters agreed or were off by only one class. An overestimation or underestimation of wetness by two classes occurred in 4% and 7% of all classification cases of group 1 and group 2, respectively. Only 0.6% (group 1) and 1.4% (group 2) of all cases were off by three wetness classes. Differences of more than three classes seldom occurred (6 of 1030). These ratings were excluded from the analysis as they occurred most likely due to assigning a class to the wrong sampling location or writing down a wrong number.

Figure 4 shows the spread in wetness class assignments among the seven different wetness classes for each sampling point. For plotting, the sampling points were sorted by the mode of the class assignments of each group for each sampling point and given ascending numbers

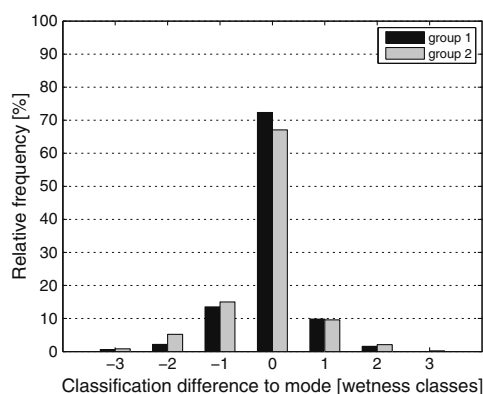


Figure 3. Deviation of wetness classifications relative to the mode of each sampling point for group 1 ( $n = 510$ ) in black and group 2 ( $n = 480$ ) in grey plotted as relative frequency distribution

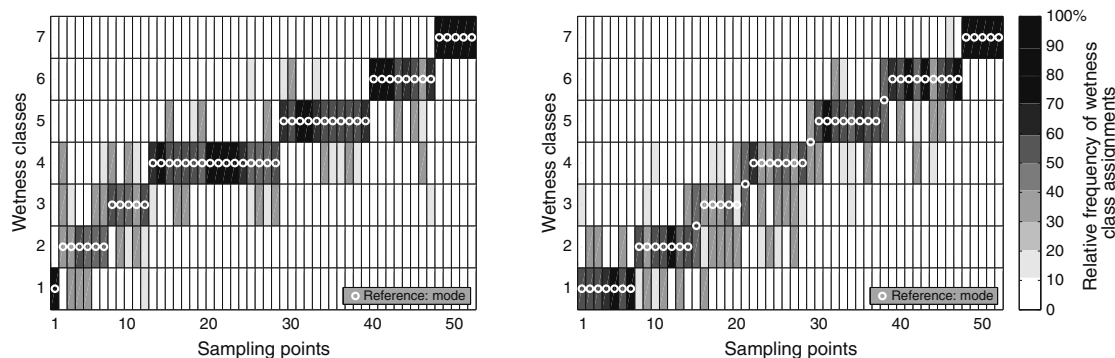


Figure 4. Wetness class assignments for each of the 52 sampling points numbered and sorted by the mode of class assignments of each group for each sampling point. Left: group 1 (morning); right: group 2 (afternoon). Grey shades indicating relative frequency of wetness class assignments. White circles show the mode as reference (in cases of two or more classes with the same frequency, the median was chosen as reference)

from 1 to 52. Grey shades indicate relative frequency of wetness class assignments and white circles show the mode that was used as reference value. Raters of both groups agreed to a large extent in wetness class assignment for wet to intermediate sampling points. This is expressed by a narrow frequency distribution (grey scale), with most or almost all assignments for a single sampling point falling within the same wetness class. A small spread of class assignments can be seen for classes 5 to 7 (for both groups) and for class 4 of group 1. Wetness classes 2 and 3 show a larger spread of class assignments. Because the raters' classification results were not known a priori, the sampling points could not be evenly distributed among the seven wetness classes. For that reason, the driest wetness class is not well represented in the dataset of group 1. Thus, the degree of agreement for the driest sampling points can only be assessed from the dataset of group 2.

To analyze if individual raters perform differently or even show a systematic bias from the reference, the mean difference of classification to the reference for all sampling points of a wetness class per rater was plotted (Figure 5). Positive differences (shown in blue) indicate a mean rater classification, which is too wet, and negative differences (shown in red) indicate a mean rater classification, which is too dry, compared with the reference. This allowed the identification of wetness classes that are difficult to assign. There was complete agreement of all raters when assigning the wettest class 7. The driest class was not well enough represented in the dataset of group 1, but from the results of group 2, it seems that some raters tended to overestimate the wetness of sampling point of the driest class. The mean differences of classification of all other wetness classes (classes 2–6) did not show a systematic pattern. Individual raters seemed to either overestimate or underestimate the wetness condition systematically, but for most of them, the mean difference was smaller than one wetness class.

To statistically quantify the degree of agreement discussed, Cohen's kappa (CK) and Krippendorff's alpha

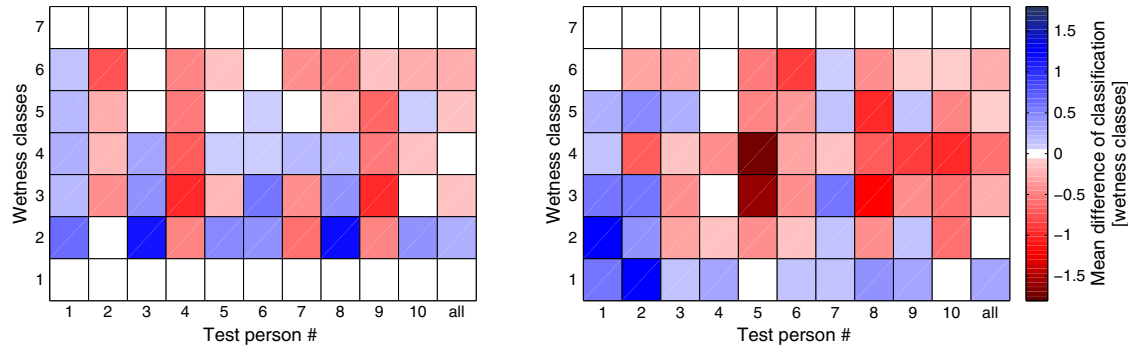


Figure 5. Mean classification difference for all sampling points of each wetness class per rater. Left: group 1 (morning); right: group 2 (afternoon). Red shades indicate mean classification of a wetness class to be too dry, blue shades to be too wet compared to the mode of each wetness class

were calculated. Both are among the few statistical measures available for assessing the agreement of different ratings within a categorical dataset. CK is used as a measure to assess concordance between two raters, or, in our case, each individual rater and a reference (Cohen, 1960). If two raters do not agree at all, CK equals zero, and if they both agree completely, CK would theoretically equal one. However, raters normally assign classes not equally frequent, which causes maximum attainable CK values ( $CK_{max}$ ) normally to be smaller than one (see Table II). CK values obtained from our dataset ranged between 0.44 and 0.79. In fact, we found CK values within 56% to 94% of  $CK_{max}$ . Measures of statistical significance of CK are seldom reported because even small kappa values can be significantly different from zero (Bakeman and Gottman, 1997). Among other reasons is prevalence, namely, the influence of classes not equiprobably assigned by two raters as well as a difference in marginal probabilities for the two raters (Sim and Wright, 2005). As common measures of statistical significance can be misleading, kappa values should be interpreted in terms of the maximum attainable kappa.

Krippendorff's alpha is a measure to assess the degree of agreement within a group of raters (Krippendorff, 2004). It is defined as one minus the ratio between observed disagreement to expected disagreement (assuming random

assignments). If the raters agree perfectly, the observed disagreement is zero and Krippendorff's alpha is one. If wetness classes would be assigned randomly, observed and expected disagreement would be equal, and Krippendorff's alpha would be zero (Krippendorff, 2011). We found a high degree of agreement with a score of 0.84 for group 1 and 0.87 for group 2.

To prove correspondence between qualitative wetness classes and quantitative measurements, the spread of volumetric water content for soil samples of each wetness class determined by the gravimetric method was plotted against the associated qualitative wetness class (Figure 6, left). In general, qualitative wetness classes did reflect differences in mean volumetric water content. However, for intermediate wetness classes, the median and the interquartile range (IQR) were not distinguishable from each other. Classes 3 and 4 were among those which were found to be most difficult to assign by the raters (see Figures 4 and 5). In particular, the IQR of class 2 spanned a large range of volumetric water content values between 40% and 75%. Wetness classes 6 and 7 were distinct from the drier ones and showed a difference in their median but their spread overlapped.

The spread of volumetric water content measured by TDR for the distributed sampling points (dataset 3) plotted against the associated qualitative wetness classes

Table II. Cohen's kappa (CK) values per rater for groups 1 and 2 for wetness classifications compared to compared with the mode of class assignments for each sample point

Rater	Group 1			Group2		
	CK	$CK_{max}$	CK as % of $CK_{max}$	CK	$CK_{max}$	CK as % of $CK_{max}$
1	<b>0.71</b>	0.88	81.1	0.50	0.77	65.0
2	<b>0.70</b>	0.86	80.9	0.49	0.71	68.6
3	<b>0.57</b>	0.83	68.5	0.73	0.95	76.4
4	<b>0.50</b>	0.77	64.4	0.71	0.87	81.8
5	<b>0.79</b>	0.83	94.3	0.50	0.77	65.1
6	<b>0.72</b>	0.90	80.2	0.55	0.82	67.0
7	<b>0.77</b>	0.82	93.9	0.73	0.91	80.1
8	<b>0.62</b>	0.85	73.0	0.44	0.78	56.5
9	<b>0.49</b>	0.79	62.3	0.66	0.86	76.4
10	<b>0.75</b>	0.93	81.3	0.71	0.86	81.7

# QUALITATIVE FIELD OBSERVATIONS OF SOIL MOISTURE PATTERNS

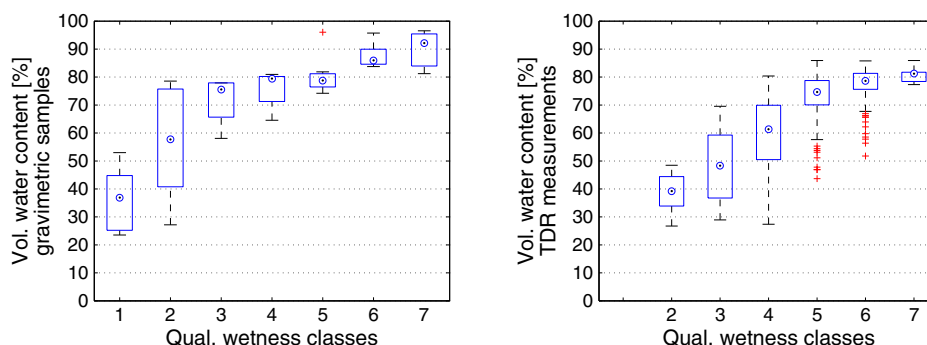


Figure 6. Left: Volumetric water content for soil samples of each wetness class determined by gravimetric measurements (dataset 2: 45 samples). Right: Volumetric water content for distributed sampling points in the catchment determined by TDR measurements (dataset 3: 454 samples). Boxes show the IQR, whiskers extend to the most extreme data point that is no more than 1.5 times the IQR from the box, outliers (red crosses) are plotted if values exceed 1.5 times the IQR and black dots represent the median

showed a similar pattern with wetter classes corresponding to higher volumetric water contents (Figure 6, right). The individual classes showed an even smaller spread (IQR) of volumetric water content compared with the gravimetric dataset 2 (Figure 6, left) but still spanned a large range. The mean values of volumetric water content of the intermediate dry and intermediate wet class (classes 2 and 3 and classes 5 and 6) corresponded well to those derived from dataset 2. Note that results for class 1 in Figure 6 (right) are not shown as no sampling point was assigned the driest wetness class in dataset 3.

## Discussion and Concluding Remarks

The overall aim of our work was to test the potential of a qualitative field method to classify soil moisture differences in wet environments and its susceptibility to subjectivity. The high level of agreement of classifications among the members of the two rater groups as well as the comparison of qualitative wetness classes to either gravimetric and TDR measurements were encouraging. In terms of subjectivity, results of the test with 20 student raters confirmed that non-experts can reliably assess soil wetness in the field. Raters showed a high level of agreement when classifying the wettest and driest classes, whereas intermediate wetness classes seemed to be more susceptible to subjective influences of individual raters and showed less agreement. Some raters showed a systematic tendency to rate specific wetness classes as too wet or too dry. Still, differences were within the range of one wetness class, and such a systematic deviation can be easily detected and corrected compared with random variations. In general, the findings apply to both groups (morning and afternoon); however, the degree of agreement had a tendency to be higher for group 1. This might be because the sampling sites had dried up between the morning and the afternoon and more sampling points fell into the intermediate range of wetness, which seemed to be more difficult to classify. The degree of agreement might be even higher if raters have some previous experience or if they get feedback after the first few ratings.

There is, however, potential to improve the method as far as differentiation of wetness classes is concerned. Because dry classes, especially wetness class 2, show considerable spread of volumetric water content (dataset 2), they could be further differentiated. Blotting paper might be one option to better distinguish between the driest classes. On the other hand, arguments could be put forward in favour of combining the three intermediate and the two wettest classes, as median and spread of volumetric water content differ little from each other. A reduction of classes, however, always affects classification results. Although fewer wetness classes would likely be easier for raters to assign, a reduction of classes would be associated with a coarser resolution of patterns and thus a loss of spatial information about soil moisture patterns. Fewer classes would also limit the identification of wetness variability over time. So we suggest applying the detailed classification with seven wetness classes, which is proposed here, leaving the option to later combine classes depending on the questions to be answered.

Limitations exist because the classification method relies on wetness indicators derived from soil surface properties. Results might be influenced by the vegetation itself, which can hold considerable amounts of water (e.g. moss) or by the litter layer being relatively dry compared to the soil layers below it. Also drizzle, dew or evapotranspiration can alter soil surface properties and therefore potentially mislead a raters' wetness classification. In terms of qualitative wetness classes and associated volumetric water content, varying soil properties (e.g. porosity) are expected to influence the relation. In our case, this was a minor issue as only gleysols can be found in the Erlenbach catchment.

Despite these issues, the qualitative method has strong advantages as it is fast to apply, needs no measuring instrument, can be applied without prior expert knowledge and is not significantly affected by subjectivity. The high degree of agreement for wet classes confirm that the qualitative method presented here is particularly useful under wet conditions. The method is not seen as a replacement but rather as a supplement to existing



quantitative measuring techniques to quickly and easily capture a large number of spatially distributed shallow soil moisture conditions over a large area of interest. Our experience revealed that one is able to assess at least five times as many locations when applying the qualitative wetness classification scheme than when using a portable TDR device (assuming five measurements per sampling location). This would allow several people, equipped with a detailed map or a GPS device to spread out simultaneously and assess hundreds to thousands of points. Training people beforehand might help to obtain an even higher degree of agreement than revealed from our test and potentially would minimize systematic bias of individuals.

There are many ways of visualizing spatio-temporal datasets: Figure 2 shows an example of how qualitative soil wetness information — here a subset of sampling points of dataset 3 — can be visualized to identify dynamics of soil moisture patterns over time (eight sampling campaigns) and identify differences in space (forested or open sampling plots). Hence, the range of variability among all sampling points or within sampling plots for a given day (columns) as well as the persistence or change over time (rows) can be made apparent in a single plot.

This method is of potential benefit for a number of applications in hydrology and other disciplines. It could help to retrieve large datasets of spatio-temporal soil moisture patterns, allowing the estimation of potential storage capacity of the topsoil, potential habitats of specific plant species or could provide additional information to constrain model parameters in the calibration procedure. For remote sensing applications these qualitative wetness classifications together with a limited number of quantitative measurements could provide ground-truthing data for a relative large spatial extent. This last example in particular indicates that supplementing existing quantitative techniques with new qualitative methods could contribute to a more efficient way by which we can assess the natural variability of soil moisture relevant for many fields of science.

## Acknowledgements

The authors thank the editor and reviewers for their useful comments as well as Tracy Ewen (Department of Geography, University of Zurich) for proofreading the manuscript. The authors also thank the master students of course GEO471, spring term 2011, for their participation in this study; J. Dusik, N. Lavanga and P. Pradhan for field and lab assistance; D. Heersink (Department of Mathematics, University of Zurich) for thoughts on the statistical analysis; T. Camek (IMKO GmbH) for TDR calibration issues; and the Swiss Federal Institute for Forest, Snow and Landscape Research WSL as well as OAK-Gemeinde Alpthal (canton Schwyz) for cooperation.

## References

- Ali GA, Roy AG, Legendre P. 2010. Spatial relationships between soil moisture patterns and topographic variables at multiple scales in a humid temperate forested catchment. *Water Resources Research* **46**(10): 1–17. DOI: 10.1029/2009WR008804
- Bakeman R, Gottman JM. 1997. *Observing interaction: An introduction to sequential analysis* 2nd edn. Cambridge University Press: Cambridge.
- Beven KJ, Kirkby MJ. 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin* **24**(1): 43–69.
- Blazkova S, Beven KJ, Kulasova A. 2002. On constraining TOPMODEL hydrograph simulations using partial saturated area information. *Hydrological Processes* **16**(2): 441–458.
- Blume T, Zehe E, Bronstert A. 2009. Use of soil moisture dynamics and patterns at different spatio-temporal scales for the investigation of subsurface flow processes. *Hydrology and Earth System Sciences* **13**: 1215–1234.
- Boelter DH. 1968. Important Physical Properties of Peat Materials. In *Proceedings of the International Peat Congress*. Quebec: Department of Energy, Mines and Resources and National Research Council of Canada, pp. 150–154.
- Bogena HR, Herbst M, Huisman JA, Rosenbaum U, Weuthen A, Vereecken H. 2010. Potential of Wireless Sensor Networks for Measuring Soil Water Content Variability. *Vadose Zone Journal* **9**(4): 1002–1013.
- Cohen J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* **20**: 37–46.
- Ellenberg H, Weber HE, Düll R, Wirth V, Werner W, Paulissen D. 1991. Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica* Vol. 18. Verlag Erich Goltze AG: Göttingen; 248.
- Feyen H, Wunderli H, Wydler H, Papritz A. 1999. A tracer experiment to study flow paths of water in a forest soil. *Journal of Hydrology* **225**(3–4): 155–167.
- Grayson RB, Western AW, Chiew FHS, Blöschl G. 1997. Preferred states in spatial soil moisture patterns: Local and non-local controls. *Water Resources* **33**(12): 2897–2908.
- Güntner A, Seibert J, Uhlenbrook S. 2004. Modelling spatial patterns of saturated areas: An evaluation of different terrain indices. *Water Resources Research* **40**(5): 1–19. DOI: 10.1029/2003WR002864
- Kollegger A. 2010. Untersuchung der räumlichen und zeitlichen Verteilung der Bodenfeuchtigkeit am Beispiel des Erlenchbaches (Alptal/Schweiz) - Vergleich einer quantitativen und einer qualitativen Messmethode. *MSc-thesis University of Zurich - Zurich*.
- Krippendorff K. 2004. *Content Analysis, an Introduction to its Methodology* 2nd edn. Sage Publications: Thousand Oaks.
- Krippendorff K. 2011. Computing Krippendorff's alpha reliability. *Department Papers (ASC)* 1–25–2011. Pennsylvania.
- McDonnell JJ, Sivapalan M, Vache K, Dunn S, Grant G, Haggerty R, Hinz C, Hooper R, Kirchner J, Roderick ML, Selker J, Weiler M. 2007. Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research* **43**(7): 1–6. DOI: 10.1029/2006WR005467
- McNamara JP, Chandler D, Seyfried M, Achet S. 2005. Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment. *Hydrological Processes* **19**(20): 4023–4038. DOI: 10.1002/hyp.5869
- Metcalfe-Smith JL. 1994. Biological Water-Quality Assessment of Rivers: Use of Macroinvertebrate Communities. In *The River Handbook*, Calow P, Petts GE, (eds). Blackwell Science Ltd: Oxford. pp. 144–171.
- Meayles E, Williams A, Ternan L, Dowd J. 2003. Runoff generation in relation to soil moisture patterns in a small Dartmoor catchment, Southwest England. *Hydrological Processes* **17**(2): 251–264. DOI: 10.1002/hyp.1122
- Michel J-C. 2010. The physical properties of peat: a key factor for modern growing media. *Mires and Peat* **6**: 1–6.
- Natural Resources Conservation Service. 1998. Estimating Soil Moisture by Feel and Appearance. *Program Aid* **1619**: 1–7.

- Petrone RM, Price JS, Carey SK, Waddington JM. 2004. Statistical characterization of the spatial variability of soil moisture in a cutover peatland. *Hydrological Processes* **18**(1): 41–52. DOI: 10.1002/hyp.1309
- Pielmeier C, Schneebeli M. 2003. Developments in the Stratigraphy of Snow. *Surveys in Geophysics* **24**(5/6): 389–416.
- Pumpanen J, Ilvesniemi, H. 2005. Calibration of time domain reflectometry for forest soil humus layers. *Boreal Environment Research* **10**: 589–595.
- De Quervain M. 1950. Die Festigkeitseigenschaften der Schneedecke und ihre Messung. *Geofisica Pura e Applicata* **18**: 179–191.
- Schleppi P, Muller N, Feyen H, Papritz A, Bucher JB, Flühler H. 1998. Nitrogen budgets of two small experimental forested catchments at Alptal, Switzerland. *Forest Ecology and Management* **101**, p. 177–185.
- Sim J, Wright C. 2005. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy* **85**(3): 257–268.
- Sivapalan M. 2005. Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale. In *Encyclopedia of Hydrological Sciences*, Anderson MG, McDonnell JJ (eds). Wiley: Chichester; 193–219.
- Thien SJ. 1979. A flow diagram for teaching texture by feel analysis. *Journal of Agronomic Education* **8**: 54–55.
- Walker JP, Willgoose GR, Kalma JD. 2001. The Nerrigundah data set: Soil moisture patterns, soil characteristics, and hydrological flux measurements. *Water Resources* **37**(11): 2653–2658. DOI: 2001WR000545
- Western A, Grayson R. 2000. Soil moisture and runoff processes at Tarrawarra. In *Spatial Patterns in Catchment Hydrology - Observations and Modelling*, Grayson R, Blöschl G (eds). Cambridge University Press: Cambridge; 209–246.
- Western A, Zhou SL, Grayson RB, McMahon TA, Blöschl G, Wilson DJ. 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal of Hydrology* **286**(1–4): 113–134.
- Western AW, Grayson, RB. 1998. The Tarrawarra data set: Soil moisture patterns, soil characteristics, and hydrological flux measurements. *Water Resources* **34**(10): 2765–2768.
- Western AW, Blöschl G, Grayson RB. 2001. Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources* **37**(1): 83–97.
- Western AW, Zhou SL, Grayson, RB, McMahon TA, Blöschl G, Wilson DJ. 2005. Reply to comment by Tromp van Meerveld and McDonnell on Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal of Hydrology* **303**(1–4): 313–315. DOI: 10.1016/j.jhydrol.2004.09.001
- Williams AG, Ternan JL, Fitzjohn C, de Alba S, Perez-Gonzalez A. 2003. Soil moisture variability and land use in a seasonally arid environment. *Hydrological Processes* **17**(2): 225–235. DOI: 10.1002/hyp.1120





## PAPER III



# Qualitative soil moisture assessment in semi-arid Africa - The role of experience and training on inter-rater reliability

M. Rinderer<sup>1\*</sup>, H. Komakech<sup>2</sup>, D. Müller<sup>1</sup>, J. Seibert<sup>1,3</sup>

<sup>1</sup> Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland; <sup>2</sup> Nelson Mandela African Institution of Science and Technology, PO Box 447 Arusha, Tanzania; <sup>3</sup> Department of Earth Sciences, Uppsala University, SE-752 36 Uppsala, Sweden

\*Corresponding author: [michael.rinderer@geo.uzh.ch](mailto:michael.rinderer@geo.uzh.ch)

---

## Abstract

Soil and water management is particularly relevant in semi-arid regions to enhance agricultural productivity. During periods of water scarcity soil moisture differences are important indicators of the soil water deficit and are traditionally used for allocating water resources among farmers of a village community. Here we present a simple, inexpensive soil wetness classification scheme based on qualitative indicators which one can see or touch on the soil surface. It incorporates the local farmers' knowledge on the best soil moisture conditions for seeding and brick making in the semi-arid environment of the study site near Arusha, Tanzania. The scheme was tested twice in 2014 with farmers, students and experts (April: 40 persons, June: 25 persons) for inter-rater reliability, bias of individuals and functional relation between qualitative and quantitative soil moisture values. During the test in April farmers assigned the same wetness class in 46% of all cases while students and experts agreed in about 60% of all cases. Students who had been trained in how to apply the method gained higher inter-rater reliability than their colleagues with only a basic introduction. When repeating the test in June, participants were given improved instructions, organized in small sub-groups, which resulted in a higher inter-rater reliability among farmers. In 66% of all classifications farmers assigned the same wetness class and the spread of class assignments was smaller. This study demonstrates that a wetness classification scheme based on qualitative indicators is a robust tool and can be applied successfully regardless of experience in crop growing and education level when an in-depth introduction and training is provided. The use of a simple and clear layout of the assessment form is important for reliable wetness class assignments.

Keywords: soil moisture, qualitative field method, inter-rater reliability, semi-arid Africa

---

## 1. Introduction

For rainfed agriculture in semi-arid regions the soil water storage is of key-importance for crop survival as it serves as the only water source during dry spells. The soil water storage is also important if water is available for irrigation. Based on differences in soil water deficits, scarce irrigation water resources can be allocated among farmers of a community in a fair manner. For farming activities like choosing the right moment to seed and for the development of crops, the moisture content in the unsaturated, shallow soil layers is of most importance.

Common techniques for measuring soil moisture are often time consuming and/or rely on expensive equipment (e.g., Time Domain Reflectometry, TDR) that needs electricity, maintenance and repair. Such instruments are also usually not available to farming communities in developing countries. Therefore local

irrigators in semi-arid Africa often visually assess the shallow soil wetness condition to decide on which plots should be allocated irrigation turns. Despite their long experience in farming, for which these leaders are respected by the community members, their assessment might be disputed. A more systematic way of soil wetness assessment based on defined criteria would relieve pressure on community leaders and assure transparency in decision making and therefore avoid conflicts among farmers.

Qualitative methods have been shown to be useful complements to quantitative measurement techniques in a number of field applications in soil science (Thien 1979), risk assessment (De Quervain 1950; cited in Pielmeier & Schneebeli 2003) and ecology (Metcalf-Smith 1994). They are based on qualitative indicators that one can identify through sight, sound or touch and that are related to quantitative properties of interest like the grain size distribution of a soil sample or the strength of a snow pack.

In hydrology qualitative indicators have been used for mapping saturated areas in some experimental

studies. Dunne & Black (1970) and Dunne et al. (1975) were the first to map saturated areas with the “Squishy Boot” method, i.e. by walking through the catchment and mapping areas with water ponding on the soil surface. Others used this method to visually identify saturated areas (Ambroise et al. 1996; Inamdar & Mitchell 2007; Latron & Gallart 2007; SNIFFER 2009). Soil hydromorphic features that are visual when digging a soil profile can be useful indicators of intermittent soil saturation (Rinderer & Seibert 2012). Also vegetation in general and individual plant species in specific can be indicators of prevailing soil moisture conditions (Ellenberg et al. 1991; Quinn et al. 1998; Kulasova et al. 2014).

The methods mentioned above do not allow different grades of soil wetness or changes in soil wetness to be captured over time. The “spade diagnosis” method, which was originally developed in the 1930s for an applied soil texture examination in the field, is one of the earliest schemes with five qualitative wetness classes (Görbing & Sekera 1947). The Natural Resources Conservation Service of the United States Department of Agriculture (1998) published guidelines for estimating soil moisture by feel and appearance for four different soil types and different soil moisture content. Blazkova et al. (2002) defined a qualitative classification scheme based on five wetness classes and used it for mapping moisture differences along transects and in a drainage ditch (for an application see also Kulasova et al. 2014). In their study, they did not utilize the full range of the five wetness classes, but aggregated the three wettest ones as they were interested in saturated areas. All these methods were not systematically tested in terms of correspondence between the qualitative indicators and the quantitative differences in soil water content and in terms of the reliability of the methods when applied by different people.

Rinderer et al. (2012) presented a soil wetness classification scheme based on characteristic, qualitative indicators for each wetness class to make class assignments more distinct. The indicators are based on the judgment of raters and include information such as whether their trousers would stay dry or get moist or wet when sitting on the ground, whether a squelchy noise could be heard, or whether water would squeeze out of the topsoil when stepping on the ground or water could be seen ponding on the soil surface. The so called “Boots & Trousers” method was tested in humid environmental conditions in terms of inter-rater reliability, influence of subjectivity and the relation between qualitative wetness classes and volumetric water content measured by the gravimetric and the TDR method.

The definitions of the three wettest classes was subsequently applied by Ali et al. (2014) to map superficial water saturation in two nested catchments in Scotland.

Despite testing the robustness of the “Boots & Trousers” method it is still not clear if this qualitative wetness classification scheme is also applicable in drier environmental conditions with different soil types. It is also unclear whether the agreement of classifications is dependent on the prior experience, the depth of the introduction or the training of the raters. We hereby define *introduction* as explanation of the method (typically 5 minutes) and *training* as practical guidance in applying the method in the field (typically 10 minutes).

In this study we present a qualitative soil wetness classification scheme that is slightly modified from the “Boots & Trousers” method (Rinderer et al. 2012), and that is capable of capturing shallow soil moisture differences in a semi-arid environment. It is adapted to the local peoples’ experience in terms of soil wetness that is optimal for seeding crops and brick making in Tanzania. The scheme is tested for its robustness and agreement between qualitative wetness classes and quantitative differences in soil water content. In particular the following questions are addressed:

- (i) Do the different qualitative wetness classes reflect actual differences in volumetric water content of the regional soil (Haplic Andosol, loamic, fluvic) of the study site?
- (ii) Does the agreement of qualitative wetness classifications depend on the participants’ experience in crop-growing or the level of education?
- (iii) Is the way in which the classification scheme is introduced to the participants and how they are trained important for achieving high agreement among raters?

## 2. Methods

### 2.1 Wetness Classification Scheme

The soil wetness classification scheme presented in this paper is based on qualitative indicators that are intuitive to local people in Tanzania from their everyday experience. In doing so, it incorporates the tacit knowledge of local peoples’ perception on soil wetness related to farming and brick making. It ranges from the driest class (#1) called “very dry – dust dry” for which one cannot see or feel any moisture in the soil at the soil surface to the intermediate class (#4), which would be the optimal

wetness for seeding plants, to the wettest class (#7) for which one could see water ponding on the soil surface (Tab. 1). The other classes represent different grades of wetness with wetness class 2 characterizing a soil sample which is dry but has some moist “look”, wetness class 3 being slightly drier than the optimal seeding conditions, wetness class 5 being optimal for making bricks and class 6 being too wet to form a brick. The indicators of the wetness scheme, namely the conditions of optimal seeding and brick making, as well as the English and Swahili class definitions were developed in the course of a field workshop and interviews with a group of local farmers.

It is not intended to tie optimal seeding conditions to a specific crop but rather to reflect farmers’ experience on good seeding conditions in general. The class “very dry – dusty dry” is also not necessarily related to the formation of a dust cloud, when stepping on the ground, as this is strongly dependent on the soil grain size distribution. It is also not intended that raters form a brick to test its stability but it is assumed that local people have good experience in imagining these conditions from their every-day life.

A vegetation cover or a litter layer as well as recent rainfall, dew or strong evaporation might affect the soil wetness conditions on the soil surface without being representative for the overall soil moisture of the soil column. To avoid these affects people were asked to always remove the upper most 5 cm of soil. It also needs to be noted that this method only assesses very shallow soil layers and not necessarily the root zone, which for some crops can be at depth of 30 to 90 cm (Weaver & Bruner 1927). However soil moisture at the surface can usually be expected to be related to soil moisture at depth for most soil types if the vertical soil moisture profile is close to equilibrium.

Icon	Class	Classname	Description
	1	very dry	"dust dry"
	2	dry	dry, but with some moist look
	3	below optimal	drier than optimal for seeding
	4	optimal	optimal for seeding crops
	5	above optimal	wetter than optimal - one can form a solid brick
	6	wet	when you step on the soil, water liquifies
	7	very wet	water ponding on the soil surface

Tab. 1.: Soil wetness classification scheme (Swahili version see supplement) with the seven wetness classes based on qualitative indicators related to best conditions for seeding and brick making.

## 2.2 Field Sites, Datasets and Test Layout

The wetness classification scheme was tested in the two farming villages Mungushi and Kichangani, in the upper Pangani basin, ca. 25 km southeast of Arusha / Tanzania (S 3° 31' 36'' / W 36° 51' 02'') (Fig. 1a). Haplic Andosols (loamic, fluviic) dominate the area where the classification scheme was tested (Fig. 2a). Soils are fertile and heavily used for growing crops, mainly beans and corn. Due to a limited amount of rainfall (below 600mm/year) (Komakech & Van der Zaag 2011) falling mainly during the rainy seasons (long rain *masika*: March – June and short rain *vuli*: October – December), agriculture in this region depends on flood irrigation

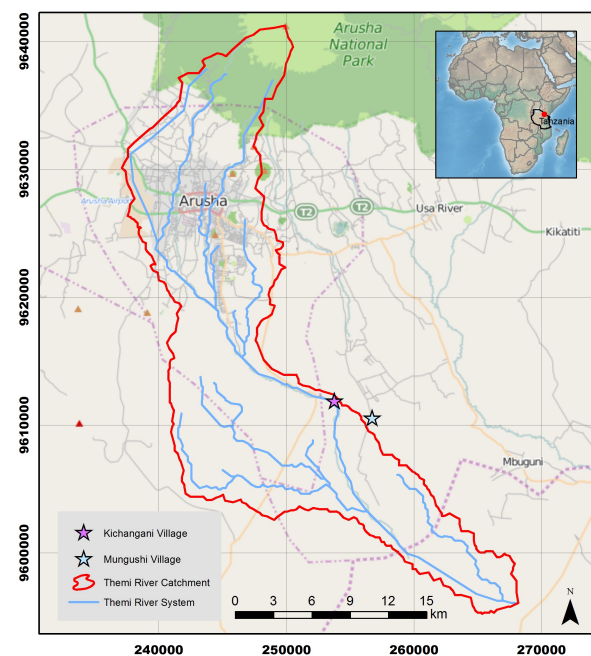


Fig. 1: Themi river catchment at Arusha / Tanzania and the two farming villages Mungushi and Kichangani were the wetness classification scheme was tested. (Background: OpenStreetMap and contributors, CC-BY-SA, insert map: Natural Earth).

during the rest of the year.

To test the wetness classification scheme we performed two experiments, one in April 2014 and another in June 2014. The first test in April was organized in the Mungushi village where 40 sampling points of different wetness were marked with flags along a 1.4 km parcours. The wetness of sequential sampling points was chosen to be random. The test involved 40 people, namely 14 farmers, 14 master students (called “students” in the following), 9 PhD students and 3 Professors. PhD students and

professors were later combined into one group called “experts”. All participants were given a brief introduction of about 5 minutes to the wetness classification scheme either in Swahili (farmers) or English (students, experts) and then were asked to individually classify the marked sites of different wetness along the parcours. Half of the farmers and students were given an additional training (~10 min) in which they were shown representative sites of wetness classes 1, 4, and 7 before the test. These two groups of participants are referred to as  $F_{trained}$  and  $S_{trained}$  in the following. Farmers and students with a basic introduction are called  $F_{basic}$  and  $S_{basic}$ , respectively. When referring to all of the farmers, students and experts we use the expressions  $F_{all}$ ,  $S_{all}$  and  $E_{all}$ . The assessment form used in April 2014 consisted of a matrix on an A4 paper (landscape format) with the number of the sampling sites appearing as rows and the wetness classes as columns (see Supplement 1 and Supplement 2). Participants were asked to tick the appropriate cell corresponding to their judgment of soil moisture conditions of a particular site.

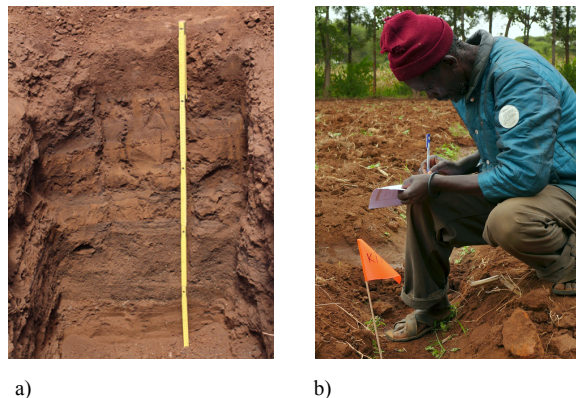


Fig. 2: a) Typical soil profile in the area where the wetness classification scheme was tested (profile depth: 1 m). b) Farmer assessing the soil wetness conditions using the qualitative soil wetness scheme. (Photo: D. Müller, M. Rinderer).

In June 2014 a similar test with 18 farmers and 7 experts was organized in the neighboring village of Kichangani (42 sampling points). The second test was intended to analyze, whether a better and longer introduction (~20 min) and training (~30 min) organized in small subgroups of 5 people and an improved layout of the assessment form, would allow farmers to gain higher inter-rater reliability than during the first test in April. The new assessment form consisted of an A4 portrait page with the class descriptions in the upper part and three columns for the soil wetness assessment (Supplement 3 and Supplement 4). The first column was pre-labeled

with “Site 1” to “Site 40” or “kituo 1” to “kituo 40” in Swahili, respectively. The second column was for the wetness class number and the third column was for optional comments. The flags, which indicated the sampling locations, were also labeled “kituo 1” to “kituo 40” to prevent potential conflicts between the number of the site and the number of wetness classes to assign. The wetness scheme remained the same except for some minor changes of class descriptions in the Swahili version.

During both tests in April and in June, volumetric water content was measured by the gravimetric method taking 100 cm<sup>3</sup> soil samples with a steel cylinder (diameter: 5 cm), at 10 cm depth below the soil surface and determining the difference in weight between the original and oven-dried sample (105 °C for 24 h).

No rainfall occurred during the day of the test in April and June and the influence of a drying up due to evaporation was considered to be small as all participants finished the test within 1 hour. In April, rainfall on the day prior to the test (no measurements available) wetted the soil while in June the fields were irrigated on the preceding days. A careful selection of sampling points was considered to guarantee the comparability between these two tests despite potential differences in infiltration patterns.

### 2.3 Statistical Analysis

To evaluate the agreement between the qualitative soil wetness classes and the quantitative measurements, the distribution of gravimetrically measured volumetric soil water content was compiled for each qualitative wetness class. To assess the agreement of qualitative wetness classifications among farmers, students and experts, the frequency distribution of classification differences relative to the median of classifications of all group members, determined at each sampling point, was analyzed. First the overall agreement among group members was investigated incorporating the classification differences of all sampling points. Furthermore the frequency distribution of wetness class assignments for each sampling point was analyzed individually in order to identify which wetness classes were distinct and which ones were more difficult to identify. The median was chosen as reference as it is a robust measure of class assignments and not affected by individual outliers.

To see if individual raters had a systematic tendency to classify some wetness classes as too wet or too dry, the mean difference of classifications to the median for all sampling points of each of the

seven wetness class was calculated for each person. Positive differences indicate a mean rater classification that was too wet and negative differences indicate a mean rater classification that was too dry compared to the reference.

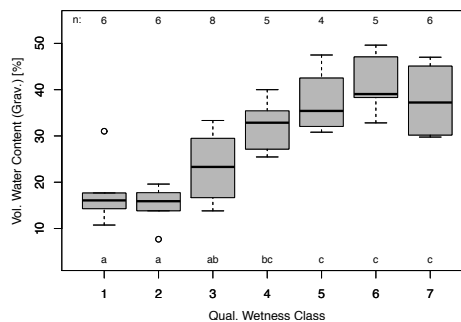
Krippendorff's Alpha (Krippendorff 2004) and Cohen's Kappa (Cohen 1960) are two statistical measures to assess the degree of agreement or inter-rater reliability among raters assigning categorical values. Krippendorff's Alpha is a measure to assess the degree of agreement within a group of raters (Krippendorff 2004). If all raters agree perfectly, the observed agreement is one and so is Krippendorff's Alpha. If wetness classes would be assigned randomly, Krippendorff's Alpha would be equal to zero as observed and expected disagreement among all raters would be equal (Krippendorff 2011).

Cohen's Kappa (CK) was used as a measure to assess concordance between two raters, or, in our case, each individual rater and a reference (Cohen 1960). If there is no agreement between the two rates other than what would be expected by chance, CK equals zero and if they both agree perfectly, CK would theoretically equal one. However, as the frequency of class assignments between two raters is normally not equal, the maximum attainable CK value ( $CK_{max}$ ) is normally smaller than one. As common measures of statistical significance can be misleading due to differences in marginal probabilities for the two raters, kappa values should be interpreted as the ratio between  $CK/CK_{max}$  (Sim & Wright 2005). In this paper, KA and  $CK/CK_{max}$  are given as percentage.

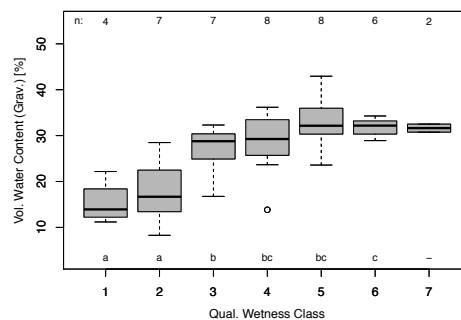
### 3. Results

#### 3.1 Qualitative and Quantitative Soil Wetness

The classes of the presented, qualitative soil wetness classification scheme reflected differences in quantitative volumetric water content of the soil samples taken during the test in April and June (Fig. 3). The median volumetric water content ranged from 16% to 39% for soil samples taken in April and from 14% to 32% for samples taken in June. The median volumetric water content and its 25%- and 75% quantiles increased for soil samples of wetness classes 2 to 6 during the test in April and for samples of classes 1 to 5 during the test in June. However soil samples of the following wetness classes had a similar median volumetric water content: classes 1 and 2; classes 6 and 7 (taken during the test in April); classes 5, 6, 7; and to a lesser extent, classes 3 and 4



a)



b)

Fig. 3: Volumetric water content for soil samples of each wetness class determined by the gravimetric method a) during test in April 2014 b) during test in June 2014. (n: sample size, letters: statistically not significantly different groups).

(taken during the test in June). A pairwise Mann-Whitney Test using an adjusted level of significance of 0.002 by Bonferroni indicated that the volumetric water content of the different qualitative wetness classes was not statistically significant. But it should be noted that the number of samples in each wetness class was low. A more relaxed significance test neglecting the Alpha-Inflation and using an unadjusted significance level of 0.05 indicated, for the test in April, that the following classes were not significantly different from each other: classes 1, 2, 3; classes 3 and 4; and classes 4, 5, 6, 7. For the dataset of the second test in June the following classes were not significantly different from each other: classes 1 and 2; classes 3, 4, 5; and classes 4, 5 and 6. Class 7 was only represented by two samples, so couldn't be assessed.

#### 3.2 Inter-Rater Reliability

In terms of the role of experience in crop growing and level of education on the agreement of wetness classifications we found that during the first test in April the  $F_{all}$  showed a lower degree of agreement

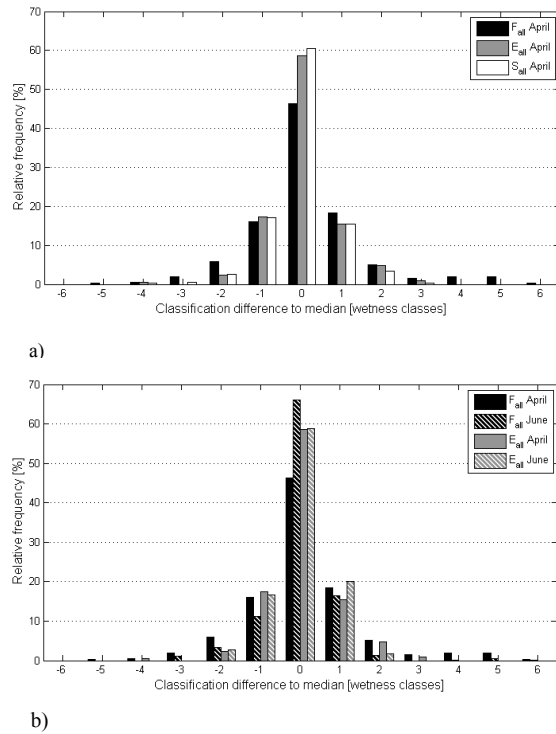


Fig. 4: Deviation of wetness class assignments a) relative to the median of all farmers ( $F_{all}$ ), all students ( $S_{all}$ ) and all experts ( $E_{all}$ ) during the test in April and b) relative to the median of  $F_{all}$  (April),  $F_{all}$  (June) and  $E_{all}$  (April),  $E_{all}$  (June).

than  $S_{all}$  and  $E_{all}$  (): In about 46% of all cases classified by  $F_{all}$  they agreed and independently assigned the same wetness class, 34% of all classifications were off the group median by one class, 11% by two classes, 4% by three classes and 5% (= 22 assignments) were off by four or more classes. In 11 times (2.5%) members of  $F_{all}$  assigned a wetness class which was off by more than four classes. The agreement of wetness classifications among  $S_{all}$  during the test in April was higher than that among  $F_{all}$  (): 60% of all cases classified by  $S_{all}$  were assigned to the same wetness class, 33% of all classification were off the group median by one class, 6% by two classes, 1% by three classes and 0.2% (= 1 assignment) were off by four classes. None of  $S_{all}$  assigned a wetness class that was off by more than four classes. The agreement of wetness classifications among  $E_{all}$  during the test in April was similar to that of  $S_{all}$  (): About 59% of all cases classified by  $E_{all}$  were assigned the same wetness, 33% of all classifications were off by one class, 7% by two classes, 1% by three classes and 0.5% (= 2 assignments) were off by four classes. No wetness classification of the  $E_{all}$  was off the group median by more than four classes.

The difference in the degree of agreement between  $F_{all}$ ,  $S_{all}$  and  $E_{all}$  during the test in April was also evident from the inter-rater reliability statistics. The Krippendorff Alpha (KA) value for  $F_{all}$  (KA: 42%) was half of KA of  $S_{all}$  (KA: 83%) and  $E_{all}$  (KA: 82%) during the test in April (Fig. 4 and Tab. 2). The median CK/CK<sub>max</sub> also differed between  $F_{all}$ ,  $S_{all}$  and  $E_{all}$  (43%, 65% and 67%, respectively; Fig. 4 and Tab. 2). The Interquartile Range (IQR) of CK/CK<sub>max</sub> was 1.8 to 3 times larger for  $F_{all}$  than for  $S_{all}$  and  $E_{all}$ , respectively (Fig. 4 and Tab. 2).

During the second test in June the agreement of class assignments among  $F_{all}$  was higher and exceeded even the agreement among  $E_{all}$  (Fig. 3): In about 66% of all cases  $F_{all}$  independently assigned the same wetness class, 28% were off the group median by one class, 4% by two classes, 1% by three classes and 1% were off by four or more classes. Only once (0.14 %) a farmer assigned a wetness class that was off by 6 classes. The agreement of wetness classifications among  $E_{all}$  was similar during the test in April and in June except that no expert was off the group median by more than two wetness classes during the second test (Fig. 3): 59% of all cases classified by  $E_{all}$  during the test in June were assigned the same wetness class, 37% of all classifications were off by one class, 4% by two classes.

Tab. 2: Inter-rater reliability statistics for the different groups (F: farmers, S: students, E: experts) during test in April and in June. ("basic" indicates only basic introduction, "trained" indicates more detailed training, "all" indicated that both subgroups have been considered). Krippendorff's Alpha and the Cohen's Kappa ratio CK/CK<sub>max</sub> can vary between 100% (perfect agreement) and 0% (no agreement other than that what would be expected by chance).

Test	Groups	Krippendorff Alpha [%]	Median CK/CK <sub>max</sub> [%] (IQR)
April	$F_{all}$	42	43 (35 - 70)
	$F_{basic}$	49	52 (46 - 59)
	$F_{trained}$	41	60 (50 - 76)
	$S_{all}$	83	65 (53 - 73)
	$S_{basic}$	81	68 (61 - 72)
	$S_{trained}$	91	83 (74 - 89)
	$E_{all}$	82	67 (58 - 70)
June	All	66	51 (34 - 62)
	$F_{all}$	76	75 (61 - 81)
	$F_{basic}$	65	75 (70 - 83)
	$F_{trained}$	87	79 (77 - 85)
	$E_{all}$	84	59 (56 - 70)
	All	78	67 (59 - 73)



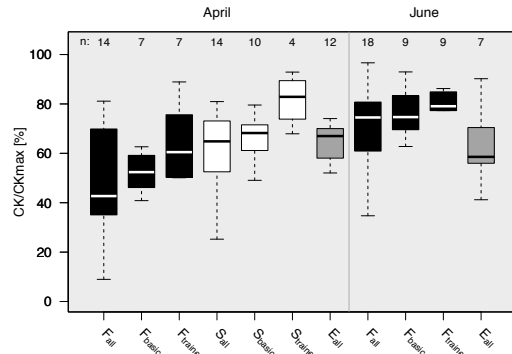


Fig. 5: Inter-rater reliability among members of individual groups tested in April and June expressed as the Cohen's Kappa ratio CK/CK<sub>max</sub> (Farmers (F): black, students (S): white, experts (E): grey; "basic" indicates the sub-group with only basic introduction, "trained" indicates the sub-group with more detailed training, "all" indicates that both subgroups have been considered; n: number of individuals in each group).

During the second test in June  $F_{all}$  achieved a similar inter-rater reliability as  $E_{all}$  (no student raters during the test in June). KA of  $F_{all}$  (KA: 76%) was more similar to KA of  $E_{all}$  (KA: 84%) and the median of CK/CK<sub>max</sub> of  $F_{all}$  (75%) even exceeded that of  $E_{all}$  (59%) during the second test in June (Fig. 4 and Tab. 2). The IQR of CK/CK<sub>max</sub> for  $F_{all}$  during the second test was almost half the IQR of the first test (Fig. 4 and Tab. 2).

In terms of the role of training on how to apply the wetness classification scheme, we found that  $S_{trained}$  during the test in April and  $F_{trained}$  during the test in June had a higher inter-rater reliability (KA and CK/CK<sub>max</sub>) compared to their colleagues with only a basic introduction (Tab. 2). The distribution of differences in classifications relative to the median of the groups was also narrower for  $S_{trained}$  during the test in April and for  $F_{trained}$  during the test in June compared to their colleagues with only a basic introduction (Fig. 3). No individual of these two groups with additional training assigned a wetness class that was off the group median by more than two classes. During the test in April the importance of additional training was not so evident among farmers. While the median CK/CK<sub>max</sub> was higher for  $F_{trained}$  compared to  $F_{basic}$ , this was not the case for KA (Tab. 2) and the spread in class assignments among  $F_{trained}$  and  $F_{basic}$  was both large. In hindsight, we partly attribute this to the use of a confusing assessment form for the test in April.

In terms of a convergence of wetness class assignments with increasing number of rated sampling points we found that during the first test in

April the median CK/CK<sub>max</sub> and KA for  $S_{all}$  and  $E_{all}$  was higher but not statistically significant for the second half of sampling points compared to the first half. This was also true for the median CK/CK<sub>max</sub> for  $E_{all}$  during the second test in June (no student raters in June).  $F_{all}$  did not have a higher median CK/CK<sub>max</sub> and KA for the second half of the sampling points compared to the first half during both tests. The median CK/CK<sub>max</sub> and KA of  $S_{trained}$  during the first test in April and  $F_{trained}$  during the second test in June was higher for the second half of the sampling points compared to the first half but the median CK/CK<sub>max</sub> of their respective colleagues with only a basic introduction was not.

### 3.3 Identifiability of Individual Wetness Classes

During the first test in April the spread of classification assignments by  $F_{all}$ ,  $S_{all}$  and  $E_{all}$  was large for all wetness classes.  $F_{all}$  had a flat frequency distribution of class assignments for all wetness classes especially for class 2 to 5 and to a lesser extent also for class 6 (Fig. 6a). Note that during both

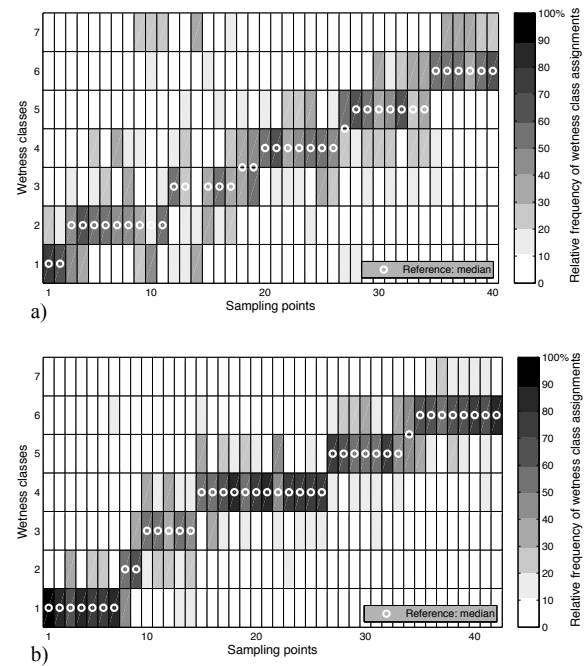


Fig. 6: Spread of classification assignments for sampling points of individual wetness classes by a) all farmers (Fall) in April and b) all farmers (Fall) in June. The difference between the two graphs shows the effect of better introduction and a clear assessment form. (grey-shades: relative frequency of wetness class assignments for each of the sampling points, white circles: median of classifications). Note that during both tests, none of the sampling points was classified as class 7 by half of Fall, and that the sampling points were distributed in random order of wetness classes in the field experiment, but were ordered here according to the median estimation for graphical clarity.

tests, half of  $F_{all}$  did not classify any of the sampling points as class 7.  $S_{all}$  and  $E_{all}$  (graphs not shown) had narrower frequency distributions of class assignments than  $F_{all}$ . The two wettest classes, class 7 and to a lesser extent class 6, showed the smallest, the dry to intermediate class 2, 3 and 4 the largest spread.

During the second test in June the spread in class assignments by  $F_{all}$  was smaller (Fig. 6b). The spread of class assignments by  $F_{all}$  improved especially for sample points of the dry to intermediate class 2 to 5 and also the second wettest class 6 between the first and the second test. The spread of class assignments by  $E_{all}$  was similar or only slightly smaller during the second test than during the first one (graphs not shown).

Regarding how training helped to better identify the wetness classes, we found that there was hardly any difference in spread of class assignments by  $F_{basic}$  and  $F_{trained}$  for the first test in April. Both groups showed large spread of class assignments for all

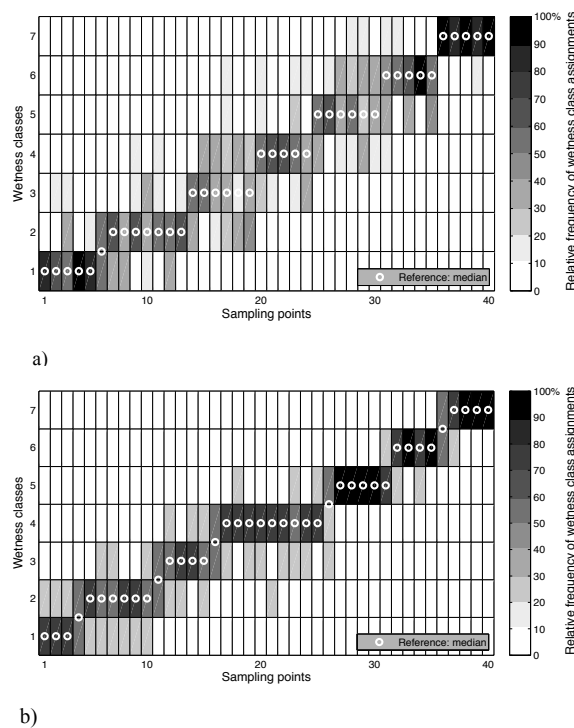


Fig. 8: Spread of classification assignments for sampling points of individual wetness classes by a)  $S_{basic}$  with basic introduction and b)  $S_{trained}$  with additional training during test in April. (grey-shades: relative frequency of wetness class assignments for each of the sampling points, white circles: median of classifications). Note that the sampling points were distributed in random order of wetness classes in the field experiment, but were ordered here according to the median estimation for graphical clarity.

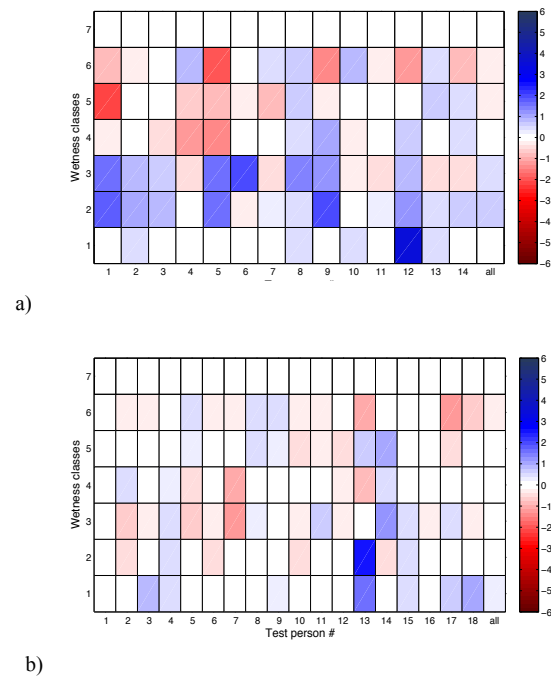


Fig. 7: Mean classification difference for all sampling points of each wetness class per test person in group  $F_{all}$  a) tested in April; b) tested in June. Red colors indicate mean classification to be too dry, blue colors to be too wet compared to the median of each wetness class.

wetness classes. In contrast,  $S_{trained}$  had narrower frequency distributions of class assignments for almost all wetness classes compared to  $S_{basic}$ , especially for the dry to intermediate classes 2 to 5 but also for the second wettest class 6 (Fig. 8). During the second test in June also the group of  $F_{trained}$  showed less spread in class assignments compared to  $F_{basic}$  (graph not shown). The improvement was noticeable for all wetness classes.

Individual people showed a systematic tendency to rate selected wetness classes either too dry or too wet. During the first test in April individual farmers as well as a few students and experts, on average showed a tendency to classify dry sampling sites too wet and to a lesser extent wet sites too dry (for  $F_{all}$  see Fig. 7a). The class 2 and 3 showed the largest mean classification differences. During the second test in June fewer individuals of farmers and experts showed a systematic bias to classify dry sites as too wet and wet sites as too dry. The mean classification difference was smaller (the whiter and pastel colors in Fig. 7b). Note that none of the sampling points had been classified as class 7 by half of  $F_{all}$  during the test in April and in June that is why the mean classification difference for this class is not given.

#### 4. Discussion

The agreement in wetness class assignments among  $S_{all}$  and  $E_{all}$  during the test in April and also  $F_{all}$  during the test in June was high which shows the robustness of the method despite being based on qualitative indicators. In 93% and 91% of all classifications the members of group  $S_{all}$  and  $E_{all}$  agreed or were off by only one wetness class during the first test in April. Despite a lower inter-rater reliability for  $F_{all}$  during the test in April, they still agreed in 81% of all cases or were off by one wetness class. These high numbers of agreement suggest that the qualitative soil wetness classification scheme in general was intuitive to local people with different levels of education and different experience in crop production.

The within-group variability of class assignments by  $F_{all}$  could be considerably reduced by a profound basic introduction organized in small subgroups, by a redesign of the assessment form layout and by a clearer labeling of the sampling sites. In 94% of all classifications the members of group  $F_{all}$  agreed or were off by only one wetness class. In June not only the site number but also the word “kituo” (English: “station”) was written on the flag. We assume that gross misclassifications of up to 6 wetness classes during the first test in April might partly be due to ticking the wrong cell of the matrix-type of assessment form. The dry to intermediate wetness classes seemed to be difficult to assign while the wettest classes were the easiest (Fig. 6). A profound basic introduction to the wetness classification scheme during the second test in June could particularly improve dry to intermediate class assignments by  $F_{all}$ . The benefit of a more detailed training was evident regardless of farming experience or education level for both,  $F_{trained}$  and  $S_{trained}$ . Not only could the within group agreement be improved but also the number of gross misclassifications of more than three wetness classes could be avoided (see Tab. 2, Fig. 3, Fig. 6, Fig. 8).

Compared to a test with master students in Switzerland (Rinderer et al. 2012), the agreement in this study was similar or lower. Classifications with an offset from the group median of more than two wetness classes were similarly frequent among Tanzanian students  $S_{all}$  (1%) and experts  $E_{all}$  (2%) compared to Swiss students (~1%), but considerably higher among Tanzanian farmers  $F_{all}$  (8%) during the first test in April. The inter-rater reliability of  $F_{all}$  (no student rates tested) during the second test in June was however similar to that of Swiss students.

A better basic introduction, organized in small sub-groups, minimized the spread of class assignments and the bias of individuals to classify wet sites as too dry and dry sites as too wet (Fig. 7). While the mean classification difference of individuals during the first test in April (see Fig. 7a) was much higher compared to the one in the study by Rinderer et al. (2012), it was similar during the second test in June (see Fig. 7b). (Note that the range of values assigned to the color ramp in Rinderer et al (2012) is different compared to Fig. 7).

The qualitative wetness classes reflected actual differences in volumetric water content of the gravimetric soil samples. However the median values of the two driest classes and the three wettest classes were very similar suggesting that a classification scheme with fewer wetness classes would be sufficient to differentiate the actual range of volumetric water content. Rinderer et al (2012) also discuss merging the two wettest classes and the three intermediate classes in their study. However a reduction of classes would be involved with a coarser resolution of the resulting patterns which might not resolve small changes in soil wetness in space and time any more. Despite being potentially less frequent, misclassification would have a larger effect on the final result when using a scheme with fewer classes.

It needs to be noted that the classification scheme by Rinderer et al (2012) was developed and tested in humid environmental conditions with moor landscapes and therefore had a different range of volumetric water content assigned to the individual wetness classes. The median volumetric water content of class 1 in the Swiss study (~38%) is similar to the median volumetric water content of class 7 (37%) in this study (Fig. 3a). This exemplifies that similar qualitative indicators on the soil surface can be associated with different volumetric water content and therefore the qualitative wetness classes need to be calibrated to the local soil types if the absolute water content is of interest.

Other limitations of this wetness classification scheme exist since only the soil surface properties are assessed, but for many crops, the soil moisture at depth is of main interest. In principle we could imagine that the classifications scheme could also be applied to a soil sample which is taken from a small pit, dug down to the depth of roots with a spade (Görbing & Sekera 1947). However digging a pit slows down the process of soil wetness assessment and soil moisture at the surface usually can be expected to be related to that at depth for most soil types if the vertical soil moisture profile is close to

equilibrium. Other potentially influencing factors are the vegetation and litter on the soil surface, wetting by dew and drizzle and drying up due to evaporation.

## 5. Conclusions

This study demonstrates the potential of a soil wetness classification scheme based on qualitative indicators that is capable of capturing shallow soil moisture differences in a semi-arid environment. It highlights the value of a detailed introduction and training to the method in gaining high agreement among individual raters but that neither experience in crop production nor a certain education level are a prerequisite for robust and comparable wetness classifications. The study also shows that the qualitative wetness classes are reflecting quantitative differences in volumetric water content.

A soil wetness classification scheme like that presented here is quick to apply, needs no expert knowledge and no measuring device, but can still provide robust and reliable results on soil moisture differences. It could be exemplified that such a qualitative method can be applied successfully in a wider range of soil- and environmental conditions (Ali et al. 2014). All these advantages make the classification scheme particularly useful and appropriate for developing countries and remote areas with limited energy supply. This method could also be used to conduct rapid spatial soil moisture assessments comprising of thousands of sampling points within a catchment. Trained farmers could send wetness classifications of their fields via SMS to a common decision support system. The spatial soil moisture patterns could then be used for model calibration and data assimilation to predict soil water stress and provide suggestions to local farmers on how to best use the available water resources. This vision of crowd-based collection of environmental data is currently under development in the project: “iMoMo - Innovative Monitoring and Modeling of Water”, funded by the Swiss Agency for Development and Cooperation (SDC) in the study area near Arusha, Tanzania.

## 6. Acknowledgments

We thank the staff and students of Nelson Mandela African Institution of Science and Technology and farmers of the Mungushi, Kichangani and Kigongoni furrow who participated in the soil moisture assessment in April and June

2014, respectively. We highly acknowledge the support of our local partners of the Pangani Basin Water Board, the Upper Kikuletwa Water Users Association (Mr. Tito Kitomari) and local village and furrow leaders. We thank Tobias Siegfried (iMoMo project coordinator), Hosea Sanga (local iMoMo-project manager), Pascal Oechslin, Beat Lüthi and Sebastian Stoll (field assistance), Guido Wiesenberg and Anett Hofmann (soil characterization), Philip Jörg (geo- and satellite data), Matthieu Bolay (anthropo-technologic issues), Alfayo Miseyeki (translation) and Tracy Ewen (proofreading the manuscript). We thank the Swiss Agency for Development and Cooperation (SDC) for financial support of the project: “Qualitative Soil Moisture Assessment in Semi-arid Conditions (Tanzania/Africa)” as part of the Global iMoMo Initiative ([www.imomohub.org](http://www.imomohub.org)).

## 7. References

- Ali, G., Birkel, C., Tetzlaff, D., Soulsby, C., McDonnell, J.J. & Tarolli, P., 2014. A comparison of wetness indices for the prediction of observed connected saturated areas under contrasting conditions. *Earth Surface Processes and Landforms*, 39(3), pp.399–413.
- Ambroise, B., Freer, J. & Beven, K., 1996. Application of a generalized TOPMODEL to the small Ringelbach catchment, Vosges, France. *Water Resources Research*, 32(7), pp.2147–2159.
- Blazkova, S., Beven, K.J. & Kulasova, A., 2002. On constraining TOPMODEL hydrograph simulations using partial saturated area information. *Hydrological Processes*, 16(2), pp.441–458.
- Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, pp.37–46.
- Dunne, T. & Black, R.D., 1970. Partial Area Contributions to Storm Runoff in a Small New England Watershed. *Water Resources Research*, 6(5), pp.1296 – 1311.
- Dunne, T., Moore, T.R. & Tayloer, C.H., 1975. Recognition and prediction of runoff producing zones in humid regions. *Hydrological Sciences Bulletin*, 20(3), pp.305–327.
- Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. & D, P., 1991. Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta geobotanica*, 18, p.248.
- Görbing, J. & Sekera, F., 1947. *Die Spatendiagnose - Ziel und Grundlage der zweckmässigen Bodenbearbeitung*, Hannover.
- Inamdar, S.P. & Mitchell, M.J., 2007. Contributions of riparian and hillslope waters to storm runoff across multiple catchments and storm events in a glaciated forested watershed. *Journal of Hydrology*, 341(1-2), pp.116–130.
- Komakech, H.C. & Van der Zaag, P., 2011. Understanding the Emergence and Functioning of River Committees in a Catchment of the Pangani Basin, Tanzania. *Water Alternatives*, 4(2), pp.197–222.
- Krippendorff, K., 2011. *Computing Krippendorff's Alpha-Reliability*, Pennsylvania.
- Krippendorff, K., 2004. *Content Analysis, an Introduction to its Methodology* 2nd ed., Thousand Oaks: Sage Publications.
- Kulasova, A., Beven, K.J., Blazkova, S.D., Rezacova, D. & Cajthaml, J., 2014. Comparison of saturated areas mapping

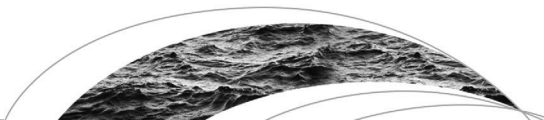
- methods in the Jizera Mountains, Czech Republic. *Journal of Hydrology and Hydromechanics*, 62(2), pp.160–168.
- Latron, J. & Gallart, F., 2007. Seasonal dynamics of runoff-contributing areas in a small mediterranean research catchment (Vallecebre, Eastern Pyrenees). *Journal of Hydrology*, 335(1-2), pp.194–206.
- Metcalf-Smith, J.L., 1994. Biological Water-Quality Assessment of Rivers : Use of Macroinvertebrate Communities. In P. Calow & G. E. Petts, eds. *The River Handbook*. Oxford: Blackwell Science Ltd., pp. 144–171.
- Natural Resources Conservation Service, 1998. Estimating Soil Moisture by Feel and Appearance. *Program Aid*, 1619, pp.1–7.
- Pielmeier, C. & Schneebeli, M., 2003. Developments in the Stratigraphy of Snow. *Surveys in Geophysics*, 24(5/6), pp.389–416.
- De Quervain, M., 1950. Die Festigkeitseigenschaften der Schneedecke und ihre Messung. *Geofisica Pura e Applicata*, 18, pp.179–191.
- Quinn, P., Ostendorf, B., Beven, K. & Tenhunen, J., 1998. Spatial and temporal predictions of soil moisture patterns and evaporative losses using TOPMODEL and the GASFLUX model for an Alaskan catchment. *Hydrology and Earth System Sciences*, 2(1), pp.51–64.
- Rinderer, M., Kollegger, A., Fischer, B.M.C., Stähli, M. & Seibert, J., 2012. Sensing with boots and trousers - qualitative field observations of shallow soil moisture patterns. *Hydrological Processes*, 26, pp.4112–4120.
- Rinderer, M. & Seibert, J., 2012. Soil Information in Hydrologic Models : Hard Data, Soft Data, and the Dialog between Experimentalists and Modelers. In H. Lin, ed. *Hydropedology - Synergetic Interaction of Soil Science and Hydrology*. Waltham: Elsevier B.V., pp. 515–536.
- Sim, J. & Wright, C., 2005. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), pp.257–268.
- SNIFFER, 2009. *A Functional Wetland Typology for Scotland - Field Survey Manual* 1st ed., Edinburgh: Scotland & Northern Ireland Forum for Environmental Research (SNIFFER).
- Thien, S.J., 1979. A flow diagram for teaching texture by feel analysis. *Journal of Agronomic Education*, 8, pp.54–55.
- Weaver, J.E. & Bruner, W.E., 1927. *Root Development of Vegetable Crops* 1st ed., New York: McGRAW-HILL BOOK COMPANY, INC.



## PAPER IV







## Water Resources Research

### RESEARCH ARTICLE

10.1002/2013WR015009

#### Key Points:

- Median groundwater levels were correlated to topographic indices
- Correlation between groundwater levels and topographic indices varied over time
- TWI assumptions were most valid during wet conditions, after peak flows

#### Correspondence to:

M. Rinderer,  
michael.rinderer@geo.uzh.ch

#### Citation:

Rinderer, M., H. J. van Meerveld, and J. Seibert (2014), Topographic controls on shallow groundwater levels in a steep, prealpine catchment: When are the TWI assumptions valid?, *Water Resour. Res.*, 50, doi:10.1002/2013WR015009.

Received 12 NOV 2013

Accepted 27 JUN 2014

Accepted article online 30 JUN 2014

## Topographic controls on shallow groundwater levels in a steep, prealpine catchment: When are the TWI assumptions valid?

M. Rinderer<sup>1</sup>, H. J. van Meerveld<sup>2</sup>, and J. Seibert<sup>1,3,4</sup>
<sup>1</sup>Department of Geography, University of Zurich, Zurich, Switzerland, <sup>2</sup>Critical Zone Hydrology Group, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, Netherlands, <sup>3</sup>Department of Earth Sciences, Uppsala University, Uppsala, Sweden, <sup>4</sup>Department of Physical Geography and Quaternary Geology, Stockholm University, Stockholm, Sweden

**Abstract** Topographic indices like the Topographic Wetness Index (TWI) have been used to predict spatial patterns of average groundwater levels and to model the dynamics of the saturated zone during events (e.g., TOPMODEL). However, the assumptions underlying the use of the TWI in hydrological models, of which the most important is that groundwater level variation can be approximated by a series of steady state situations, are rarely tested. It is also not clear how well findings from existing hillslope studies on sites with transmissive soil can be transferred to entire catchments with less permeable soils. This study, therefore, evaluated the suitability of selected topographic indices to describe spatial groundwater level variations based on time series from 51 groundwater wells in a 20 ha catchment with low-permeability soils in Switzerland. Results showed that median groundwater levels were correlated to slope, curvature, and TWI, but the strength of correlation depended on whether the indices characterized the local topography or the topography of the upslope contributing area. The correlation between TWI and groundwater levels was not constant over time but decreased at the beginning of rainfall events, indicating large spatial differences in groundwater responses, and increased after peak flow, when groundwater levels could be considered to be spatially in a steady state. Our findings indicate that topographic indices are useful to predict median groundwater levels in catchments with low-permeability soils and that the TWI assumptions are best met when groundwater levels change slowly.

### 1. Introduction

The spatio-temporal variation in groundwater levels and, thus, groundwater storage within a catchment significantly influences catchment runoff response [McGlynn *et al.*, 2004; Zehe *et al.*, 2005; Spence *et al.*, 2009]. Temporal differences in the area where groundwater storage capacity is exceeded as a result of rainfall or snowmelt govern the changing patterns of runoff source areas and overland flow connectivity, as described in the variable source area concept [Hewlett and Hibbert, 1967; Ambrose, 2004; Gomi *et al.*, 2008]. Similarly, perched groundwater levels on hillslopes determine the activation of subsurface flow pathways. When these pathways become hydrologically connected to the stream network, this can result in a rapid increase in runoff [Spence and Woo, 2006; Tromp-van Meerveld and McDonnell, 2006; Laudon *et al.*, 2007; Lehmann *et al.*, 2007]. While temporal differences in groundwater levels are important for understanding runoff processes during rainfall events, average groundwater conditions can serve as an indicator of typical wetness conditions in a catchment and its average storage capacity. As continuous groundwater level monitoring is restricted to selected sites, understanding the processes and controlling factors that lead to spatial variability in groundwater levels in a catchment is important. Quantifying relations between groundwater levels and site characteristics, such as the topographic characteristics of the monitoring site and its upslope contributing area, soil and bedrock properties, and vegetation, enables the prediction of groundwater levels at unmonitored sites. This is a prerequisite for identifying spatial patterns of groundwater above an impeding soil or bedrock layer and its spatial connection, especially to the stream network.

Several studies have demonstrated that surface and subsurface topography, vegetation, soil- and bedrock properties control the spatial variability in groundwater levels. As groundwater levels are the result of local

drainage, local recharge from infiltration, and groundwater input from upslope, it is necessary to differentiate between characteristics of the monitoring site itself (local controls) and those that are representative of the upslope contributing area (upslope controls). In mountain catchments with often shallow soils and groundwater tables, topography is assumed to be a major driver of spatial differences in groundwater levels as the gravitational potential is a dominant part of the total potential [Anderson and Burt, 1978]. This important role of topography was recognized early and forms the basis of several conceptual hydrological models, such as TOPMODEL [Beven and Kirkby, 1979] and TOPOG [O'Loughlin, 1986].

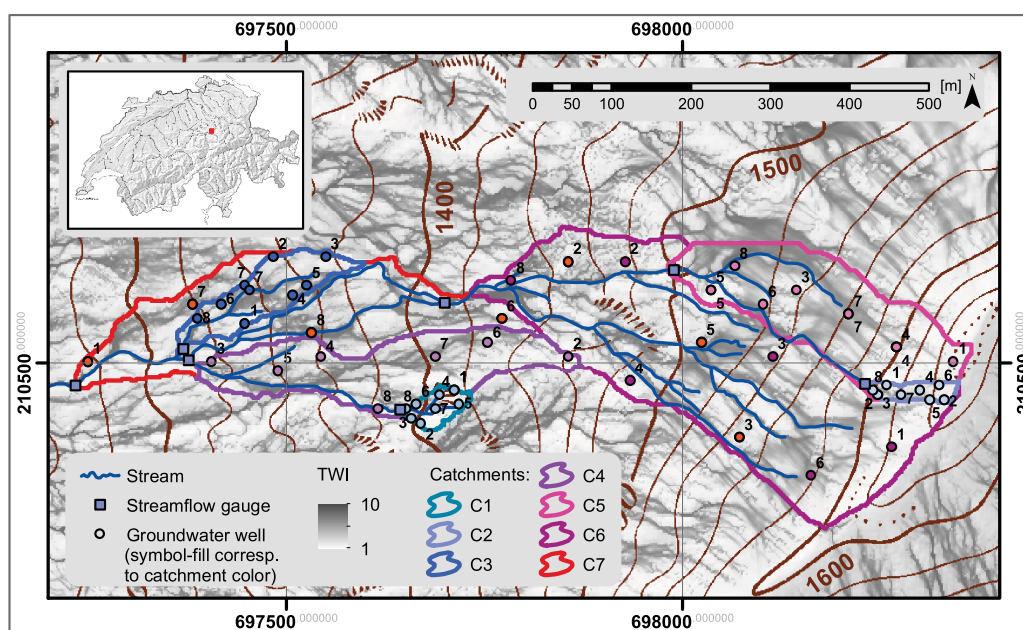
Many of the topography-based, hydrological models assume that sites with the same Topographic Wetness Index (TWI;  $\ln(a/\tan\beta)$ , where  $a$  is the upslope contributing area per unit contour length (m) and  $\beta$  is the local slope ( $^\circ$ )), have a similar groundwater response. This consideration is based on the assumptions that the local slope is a proxy of the local hydraulic gradient and that the whole upslope contributing area contributes to groundwater flow toward the site [Beven and Kirkby, 1979]. Furthermore, it is assumed that spatial groundwater table variations can be approximated by successions of steady state situations, implying for each point in time an equilibrium between inflow from the upslope contributing area and local drainage everywhere in the catchment. This implies a spatially persistent pattern of groundwater levels in a catchment. In the following, we refer to these assumptions as the *TWI assumptions*.

With the growing popularity of the TOPMODEL concept in the 1980s and 1990s, a series of studies investigated the relations between topographic indices, especially the TWI, and groundwater levels. A good agreement was found in some studies, mainly during wet conditions [Anderson and Burt, 1978; Burt and Butcher, 1985] and for sites with shallow groundwater tables [Troch et al., 1993], whereas other studies reported poorer agreements, which could partly be attributed to flat terrain [Barling et al., 1994] or transmissive soils [Seibert et al., 1997]. Some studies restricted monitoring to near-stream and footslope locations and measured groundwater levels at a coarse temporal resolution, which may also have contributed to the contradictory findings [Moore and Thompson, 1996; Buttle et al., 2001].

Distinct differences in the groundwater response have been observed for wells in the riparian zone and the upper hillslope zone [Seibert et al., 2003; Haught and van Meerveld, 2011]. While water levels in riparian wells were well correlated with streamflow in these studies, they were not for the upland sites. In other studies, water levels increased earlier in upland wells than in footslope sites due to differences in surface and bedrock topography or soil depth [Tromp-van Meerveld and McDonnell, 2006; Rodhe and Seibert, 2011; Penna et al., 2014]. These differences in groundwater response might partly explain why modeled groundwater levels did not agree with observations, when using TWI-based models or TWI as an external drift function for interpolating groundwater table elevations [Seibert et al., 1997; Desbarats et al., 2002].

The site characteristics that are most strongly correlated to groundwater levels and therefore are considered to control groundwater levels have been investigated only in a few studies. Individual Spearman rank correlation analysis showed that mean relative groundwater levels were correlated to land use classes, soil properties, local slope, hillslope position, and well depth, but not upslope contributing area, local plan and local profile curvature, saturated hydraulic conductivity, and vegetation properties for hillslopes in southern Germany with sandy loam textured soils [Bachmair and Weiler, 2012]. However, when applying a nonparametric multivariate technique (random forest approach [Breiman, 2001]) to predict the mean relative groundwater levels using the same independent variables as listed above, saturated hydraulic conductivity and local profile curvature were the most importance predictors, followed by topographic variables such as local slope, local plan curvature, and upslope contributing area. The explained variance of mean relative groundwater levels using the random forest approach was only 30%.

Bachmair and Weiler [2012] present the only study that reported seasonal differences in the importance of site characteristics on groundwater levels. They found that correlations between mean water tables and site characteristics were lower during summer than during fall, winter and spring. Correlations between mean groundwater levels with site characteristics were even lower for individual events [Bachmair and Weiler, 2012]. We are not aware of any previous study that investigated the change in correlation between groundwater levels and site characteristics during events. This is maybe partly because, until recently, continuous measurements of groundwater levels at many points in a catchment were not feasible. As the groundwater response is known to vary throughout a catchment during a rainfall event and patterns in groundwater



**Figure 1.** The study catchment showing the seven nested subcatchments with a streamflow gauging station at each outlet, the location of the spatially distributed groundwater wells (numbered from 1 to 8 in the order of increasing wetness in each subcatchment), and Topographic Wetness Index in the background (topographic map: reproduced under permission of swisstopo (BA12059)).

levels, therefore, change over time, a temporal change in the correlation between groundwater levels and topographic indices is likely.

Previous studies have investigated groundwater dynamics and key controls mainly on hillslopes or at the riparian-hillslope interface, but less is known about catchment-wide variability in groundwater levels. Furthermore, most of the previous studies have been conducted at sites with transmissive soils [Seibert *et al.*, 2003; Tromp-van Meerveld and McDonnell, 2006; Detty and McGuire, 2010]. The dominant processes and catchment characteristics (e.g., soil properties, topography) that determine groundwater dynamics are expected to be different in catchments with less permeable soils (e.g., Gleysols) because groundwater levels are expected to be more persistent, quicker to respond, and more frequent, because of the lower storage deficit and smaller drainable porosity compared to catchments with transmissive soils.

This study, therefore, aimed to assess the influence of topographic characteristics on groundwater levels in a steep headwater catchment with low-permeability soils by addressing the following questions:

1. To what extent does topography control median groundwater levels in a catchment with low-permeability soils?
2. Are there differences in the correlation of median groundwater levels with local and upslope topographic characteristics?
3. Does the correlation between topography and groundwater levels vary over time?

## 2. Methods

### 2.1. Site Description

The 20 ha headwater study catchment is located in the Alptal, a pre-alpine valley about 40 km southeast of Zurich, Switzerland (Figure 1). The Alptal region and particularly the Erlenbach catchment is known for a long history of research on the influence of forests on runoff, water quality, and bedload transport [Hegg *et al.*, 2006]. However, the Erlenbach catchment was not chosen for this study because it is affected by anthropogenic drainage. Instead a 20 ha neighboring catchment was investigated. Mean annual precipitation is 2300 mm/yr, of which about 30% falls as snow, and is evenly distributed throughout the year [Feyen *et al.*, 1999]. The catchment extends from 1270 m asl. to 1650 m asl. and has an average slope of 35%.

Landslides and soil creep have developed a sequence of steeper and flatter landscape units, each with complex microtopography, and a dense natural drainage network (205 m/ha) with most channels not being deeply incised, except for the main channel close to the catchment outlet. Moor landscapes and wet grassland areas have formed in flat or concave parts of the catchment (ca. 7 ha), while steeper slopes and ridge sites have open coniferous forest stands (*Picea abies* L. with an understory of *Vaccinium* sp.) [Hagedorn et al., 2000] (ca. 11 ha). Parts of the upper catchment (ca. 2 ha) are used for cattle grazing during summer (area based on aerial photographs from 2007). The spatial distribution of soil types and soil depth are related to differences in local topography. In wet depressions (mainly grassland), where the water table is persistently close to the soil surface, a mollic Gleysol with a topsoil high in carbonate can be found. The mineral soil consists of a permanently reduced Bg horizon, with typically 43% clay, 42% silt, and 15% sand [Schleppi et al., 1998]. At the ridge sites, where the water table is normally more than 40 cm below the soil surface, trees grow on an umbric Gleysol with an oxidized Bw horizon (49% clay, 46% silt, and 5% sand) [Schleppi et al., 1998; Hagedorn et al., 2001] with macropores. Soil depth varies between 0.5 m at ridges to more than 2.5 m in depressions. The bedrock consists of a poorly permeable clay-rich Flysch with calcareous sandstone and argillite and bentonite schist layers [Mohn et al., 2000].

## 2.2. Monitoring Network and Measurements

The study catchment consists of seven nested subcatchments (C1–C7) of varying size ( $\sim 0.2$ ,  $\sim 1$ ,  $\sim 3.5$ ,  $\sim 12$  to 20 ha; see Figure 1). In contrast to most previous studies, where groundwater levels were measured along transects or on a single hillslope, the monitoring network of this study was designed to provide a good spatial coverage and to capture wet and dry sites within each subcatchment. As field observations suggested that TWI might be a good indicator of soil wetness, TWI (calculation described in section 2.3) was used to determine the locations of the monitoring sites. For each subcatchment, the pixels were grouped into eight TWI classes with equal frequency. The coordinates of the monitoring sites were determined by selecting the pixels with a TWI similar to the median TWI of each class. As the subcatchments were nested, five monitoring sites overlapped, resulting in 51 monitoring sites with continuous groundwater level observations (Figure 1). The monitoring sites included 8 ridge site, 22 midslope, and 21 footslope or depression sites. Of the 51 monitoring sites, 25 had a mollic Gleysol and 26 had an umbric Gleysol profile; 20 sites were forested and 31 were located in grassland. Soil depth was not statistically significantly different between mollic and umbric Gleysol monitoring locations (Mann-Whitney  $U = 288$ ,  $p = 0.5$ ). Soil depth was correlated to the local slope (Spearman rank correlation coefficient  $r_s = -0.44$ ,  $p = 0.001$ ).

All boreholes were hand-augered down to the parent material. The mean depth was 1.06 m (min: 0.46 m, max: 2.16 m). The wells consist of a PVC pipe of 4 cm diameter, screened over the full length up to 10 cm below the surface; the borehole was backfilled with coarse filter sand after installation of the pipe. To prevent water entering the well and auger hole from the soil surface, the filter pack was sealed with bentonite and plastic foil 5–10 cm below the soil surface. Water levels were measured in the wells between September 2010 and November 2012 using Odyssey capacitance water level loggers (Dataflow Systems Pty Limited). The measurement interval was 5 min during summer (May until December) and 10 min during winter. Groundwater level data were checked with manual water level measurements when downloading the data, every 2–3 months. Saturated hydraulic conductivity of the mineral soil layer was determined by the Bouwer and Rice [1976] method based on at least three slug and bail tests at each groundwater-monitoring site during summer 2012.

Stream stage was measured every 5 min at each of the seven subcatchments during summer (May until December) 2011 and 2012 using pressure loggers (DL/N 70 by STS, Sensor Technik Sirmach AG) and every 10 min during winter 2011 and 2012 using capacitance water level loggers (Odyssey). HS flumes (subcatchment C1 and C2) and 90° V notch weirs (subcatchments C3, C4, and C5) were used in channels with moderate sediment transport. Stage was converted into streamflow using rating curves [U.S. Department of the Interior, 2001] that were checked by repeated salt dilution measurements during seven events of different magnitude and a low-flow period. For the largest and second largest catchments (C6 and C7), stage was recorded in a natural cross section as weir construction was not possible. Changes in the natural cross section were documented monthly and deemed to be minor for the study period. Salt dilution was used to determine the rating curves for these cross sections.

Precipitation, air temperature, and barometric pressure were measured at a permanent meteorological weather station 1 km from the experimental catchment at 1219 m asl. Precipitation and air temperature

**Table 1.** Spearman Rank Correlation Coefficients ( $r_s$ ) for the Relations Between Median (Relative and Absolute) Groundwater (GW) Levels and Selected Local and Upslope Topographic Site Characteristics<sup>a</sup>

Site Characteristic	Method/Reference	Units	Type	$r_s$ (Median Relative GW Level)	$r_s$ (Median Absolute GW Level)
Local slope	Calculated based on the D $\infty$ flow algorithm [Tarborton, 1997]	%	Local	<b>−0.67</b>	<b>−0.57</b>
Mean slope of upslope contribution area	Upslope contributing area determined by the MD $\infty$ flow algorithm [Seibert and McGlynn, 2007]	%	Upslope	−0.23	−0.15
Local curvature	Second derivative of a bivariate quadratic surface through a local 3 × 3 kernel [Travis et al., 1975; Evans, 1980]	-	Local	−0.23	−0.26
Mean curvature of upslope contribution area	Upslope contributing area determined by the MD $\infty$ flow algorithm [Seibert and McGlynn, 2007]	-	Upslope	<b>−0.80</b>	<b>−0.77</b>
Upslope contributing area	Determined by the MD $\infty$ method [Seibert and McGlynn, 2007]	m <sup>2</sup>	Upslope	<b>0.69</b>	<b>0.70</b>
Topographic wetness index (TWI)	ln(a/tan $\beta$ ) [Beven and Kirkby, 1979]	ln(m)	Upslope	<b>0.78</b>	<b>0.77</b>
Mean TWI of upslope contribution area	Upslope contributing area determined by the MD $\infty$ flow algorithm [Seibert and McGlynn, 2007]	ln(m)	Upslope	<b>0.62</b>	<b>0.61</b>

<sup>a</sup>Bold:  $r_s$  statistically significant with  $p < 0.05$ ; data: September 2010 to November 2012.

were measured every 10 min, while barometric pressure was measured every 5 min. There is no reliable information on the spatial patterns of precipitation in the catchment, but we expect the altitudinal gradient in precipitation to be small and differences in the timing of the onset of precipitation to even out over the study period of 27 months.

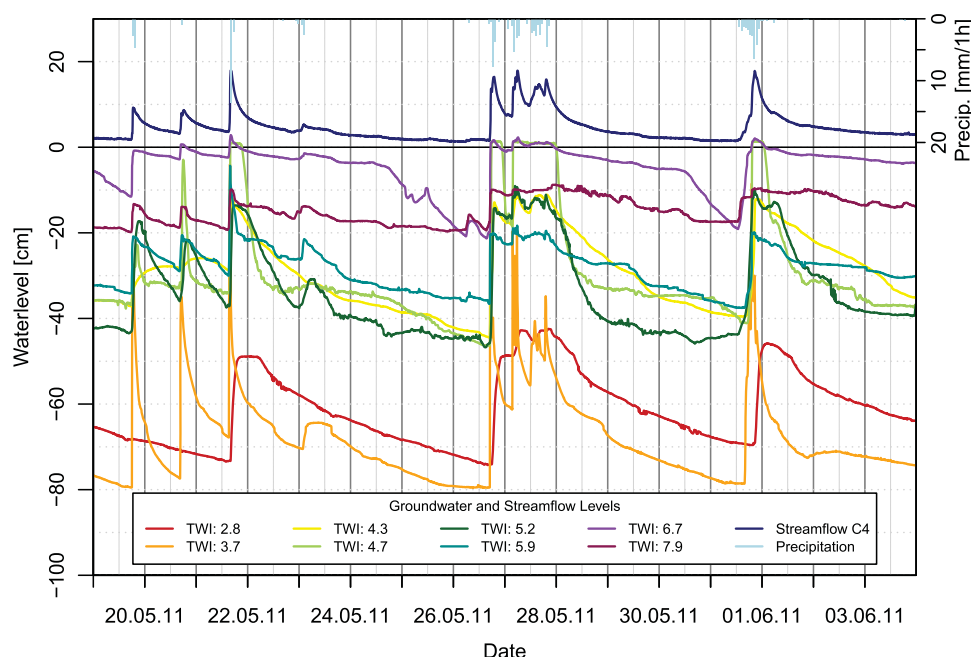
### 2.3. Site Characteristics

We defined *key controls* as the characteristics that are significantly correlated to the median values of the groundwater level time series of all sites and therefore can explain parts of the observed spatial variability in median groundwater levels across the catchment. As we expected differences in the importance of local characteristics of a site and the characteristics of its upslope contributing area, we defined *local controls* as properties that characterize the monitoring site itself and *upslope controls* as the properties that characterize the upslope contributing area. The site characteristics selected for this study were local slope, local curvature, TWI, upslope contributing area, mean slope, mean curvature, and mean TWI of the upslope contributing area (Table 1).

The topographic site characteristics were calculated based on a Digital Terrain Model (DTM) derived from LiDAR. DEM resolutions of 2, 4, 6, 8, and 10 m were tested, and 6 m was found to be the optimum for capturing the prominent morphologic features (ridges and depressions) without being obscured by microtopography. For all upslope characteristics, the triangular multiple flow direction algorithm [Seibert and McGlynn, 2007] was used for downslope routing of the accumulated area. All indices were calculated using the open source software SAGA-GIS [Conrad, 2007]. The mean values of the tested topographic indices for the upslope contributing area might not be representative if they are based only on a few pixels, but we consider this effect to be minor because the correlations between upslope controls and median groundwater levels were similar when sites with an upslope contributing area smaller than 125 m<sup>2</sup> (lower 25% quantile; equivalent to ca. 3 pixels) were excluded.

### 2.4. Analytical Methods

To quantify the relation between the topographic characteristics and groundwater levels, the Spearman rank correlation coefficient ( $r_s$ ) was determined [Spearman, 1904]. For characterizing the average system state, we chose median instead of mean groundwater levels since these are less influenced by extremes and more robust for censored data (i.e., when the groundwater level falls below the bottom of the groundwater well). As soil depth and, thus, well depth differed between sites, groundwater levels were scaled by the soil depth (1 = water level at the soil surface, 0 = dry well). We refer to these scaled water levels as *relative groundwater levels* throughout the remainder of this text, whereas the unscaled water levels are referred



**Figure 2.** Groundwater dynamics of eight, representative wells with a different Topographic Wetness Index (TWI) showing distinct differences in response, peak groundwater level and recession, together with stream stage in catchment C4. Note: groundwater levels are given in cm (negative values indicate distance below the surface).

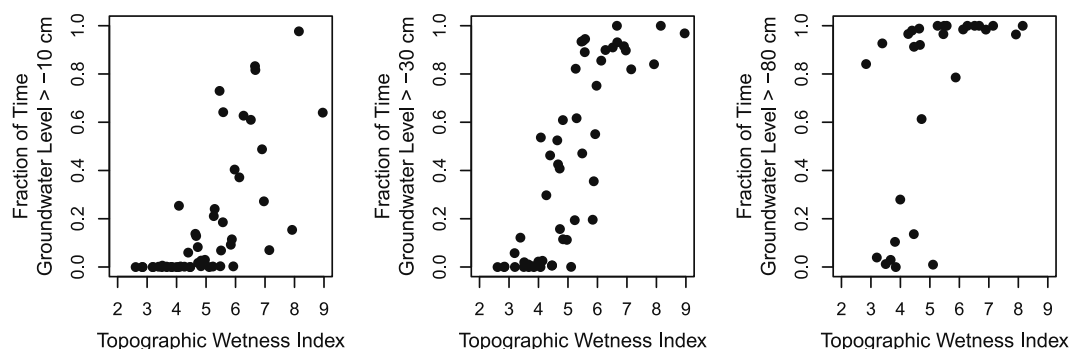
to as *absolute groundwater levels*. The software R (version 2.14.1) was used to analyze the data [R Development Core Team, 2005]. A statistical significance level of  $\alpha = 0.05$  was applied throughout the study.

The continuous measurement of groundwater levels also allowed the investigation of the temporal variation in the correlation between groundwater levels and TWI. Groundwater levels and streamflow data were aggregated to hourly time steps by calculating the mean to remove noise in the data. Rank correlation coefficients between groundwater levels and TWI were then calculated for each hour and related to streamflow and the relative change in streamflow ( $dQ/Q$ ) in subcatchment C5. Streamflow was assumed to be an indicator of the system state, and C5 was used because it provided the most complete runoff series. Data points were further classified according to three hydrologically relevant seasons in the Alptal region: the *growing season* from the beginning of June until the end of September with frequent rainfall events, the *dormant season* between the beginning of October and the end of January and *spring*, including snowmelt between the beginning of February and the end of May.

### 3. Results

#### 3.1. Characterizing Groundwater Variability

Groundwater dynamics varied spatially across the small mountain headwater catchment (Figure 2). Most sites with a  $TWI < 4$  did not respond during every rainfall event and seemed to have a threshold type of response behavior. Most other sites responded during the majority of the rainfall events but differed in their peak groundwater level. There was also a difference in the lag time between the start of a rainfall event and the rise of the groundwater level. For some sites, the recession limb was almost as steep as the rising limb, while for others, it took several days to return to the base level. Even for sites with a similar rate of recession, the median absolute groundwater level was distinctly different. Despite differences in the groundwater hydrographs of individual sites, similarities could be identified. Sites with a  $TWI > 6$  responded faster than sites with a  $TWI < 4$ . For most sites with a  $TWI > 6$ , the groundwater levels were close to the soil surface most of the time, the rise during an event was small compared to other sites, and the groundwater levels tended to remain elevated for several days after rainfall events. Sites with a  $TWI$  between 4 and 6 showed the highest response frequency and amplitude and differed most in mean groundwater levels, while sites with a  $TWI < 4$  only responded to large rainfall events or events with a high rainfall intensity.



**Figure 3.** Fraction of time the groundwater level at each site was within (left) 10 cm, (middle) 30 cm, or (right) 80 cm from the soil surface as a function of Topographic Wetness Index (TWI,  $\ln(m)$ ).

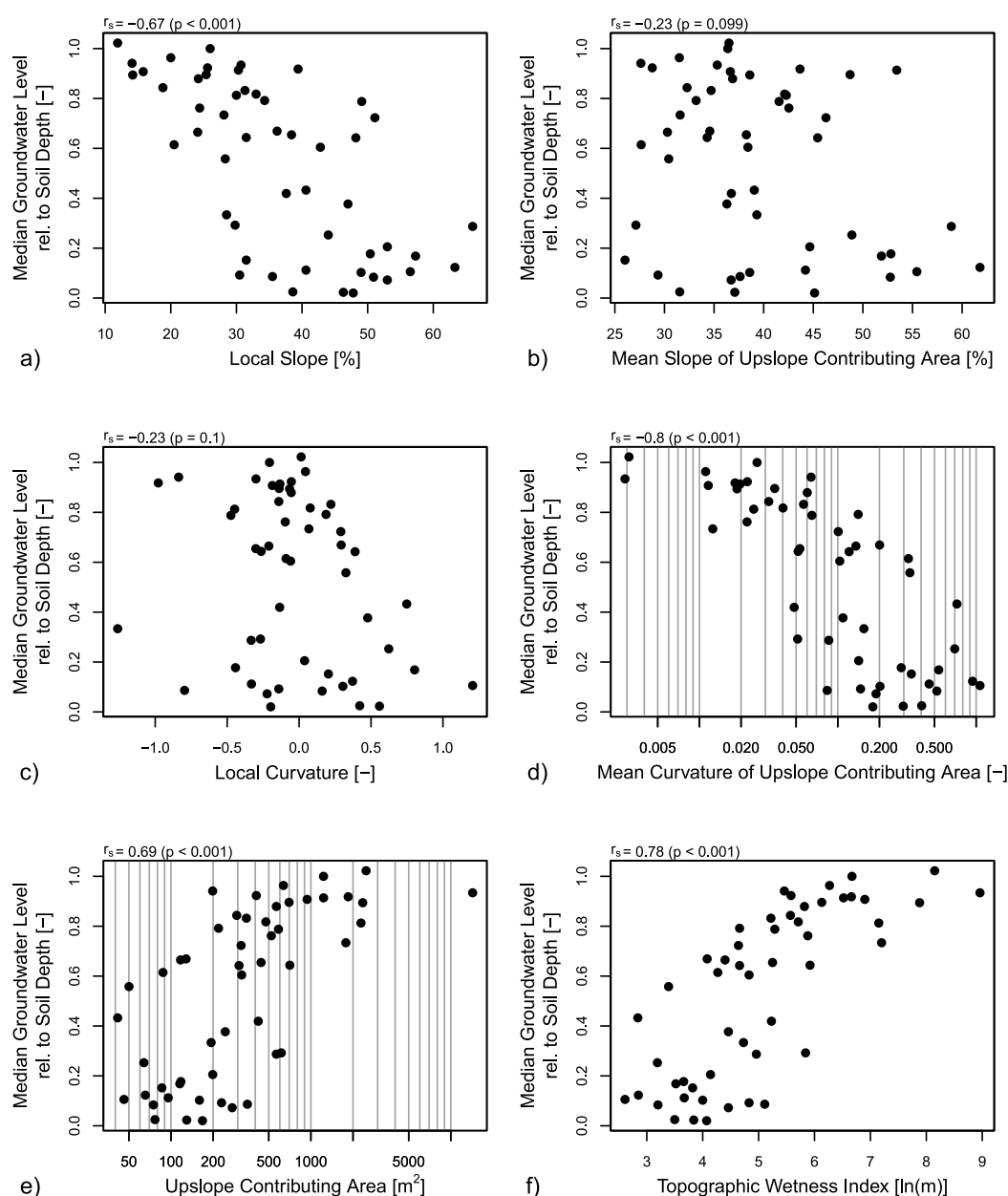
The skewness of the frequency distributions of the groundwater levels characterizes the water table dynamics at each site. Sites with a  $TWI < 4$  had predominantly positively skewed frequency distributions (i.e., mainly low water levels), while sites with a  $TWI > 6$  were predominantly negatively skewed (i.e., mainly high water levels). The groundwater level distributions of sites with a local slope  $< 30\%$  were predominantly negatively skewed, while for sites with a local slope  $> 50\%$  they were predominantly positively skewed. The skewness of the groundwater frequency distribution was correlated to all topographic indices considered in this study (e.g., local slope:  $r_s = 0.68$ , TWI:  $r_s = -0.69$ ), except local curvature. The fraction of time the wells were filled to a certain level below the soil surface was also related to topography. For sites with a  $TWI < 4$ , groundwater levels were almost never within 10 cm from the surface, whereas for sites with a  $TWI > 4$  there was considerable spread and a weak tendency of an increasing fraction of time with water levels within 10 cm from the soil surface, with increasing TWI (Figure 3, left). This relation was more pronounced when analyzing the fraction of time that water levels were within 30 cm from the soil surface, especially for sites with a TWI between 4 and 6 (Figure 3, middle). Only sites with a  $TWI > 7$  almost always had a water level within 30 cm from the surface. A similar pattern could be observed for the fraction of time water levels were within 50 cm from the soil surface (not shown). All sites had a water level within 80 cm from the surface for  $> 80\%$  of the time, except for nine sites with a  $TWI < 5$  (Figure 3, right).

### 3.2. Correlation Analysis

The median relative groundwater levels were correlated to most of the selected topographic indices. However, the strength of the correlation differed for the local and upslope topographic characteristics. The median groundwater levels were correlated to the local slope ( $r_s = -0.67$ ) but not to the mean slope of the upslope contributing area (Figures 4a and 4b and Table 1). Steeper sites generally had lower median relative groundwater levels. While sites with a local slope between 30 and 50% had median relative groundwater levels over almost the entire range (between 0.05 and 0.9), flatter and steeper sites had median relative groundwater levels  $> 0.6$  and  $< 0.3$ , respectively. These results were similar for the median absolute groundwater levels, but the correlation coefficients were lower (Table 1).

In contrast to slope, the median relative groundwater levels were highly correlated to the mean curvature of the upslope contributing area ( $r_s = -0.80$ ) but not to the local curvature (Figures 4c and 4d and Table 1). Most sites had a local curvature between  $-0.5$  and  $0.5$ , but regardless of being convex or concave, the median relative groundwater levels ranged between 0 and 1 (Figure 4c). The correlations were similar for the median absolute groundwater levels (Table 1).

The median relative groundwater levels were also correlated to the upslope contributing area ( $r_s = 0.69$ ) (see Figures 4e and Table 1). For the majority of sites with an upslope contributing area smaller than about  $200 \text{ m}^2$ , the median relative groundwater level was less than 0.3, except for five sites that had median relative groundwater levels between 0.4 and 0.7. For sites with an upslope contributing area between 200 and  $600 \text{ m}^2$ , the median relative groundwater levels varied over the entire range. For sites with an upslope contributing area larger than  $600 \text{ m}^2$ , median relative groundwater levels were higher than 0.7. The upslope contributing area was similarly correlated to the median absolute groundwater levels ( $r_s = 0.70$ ).

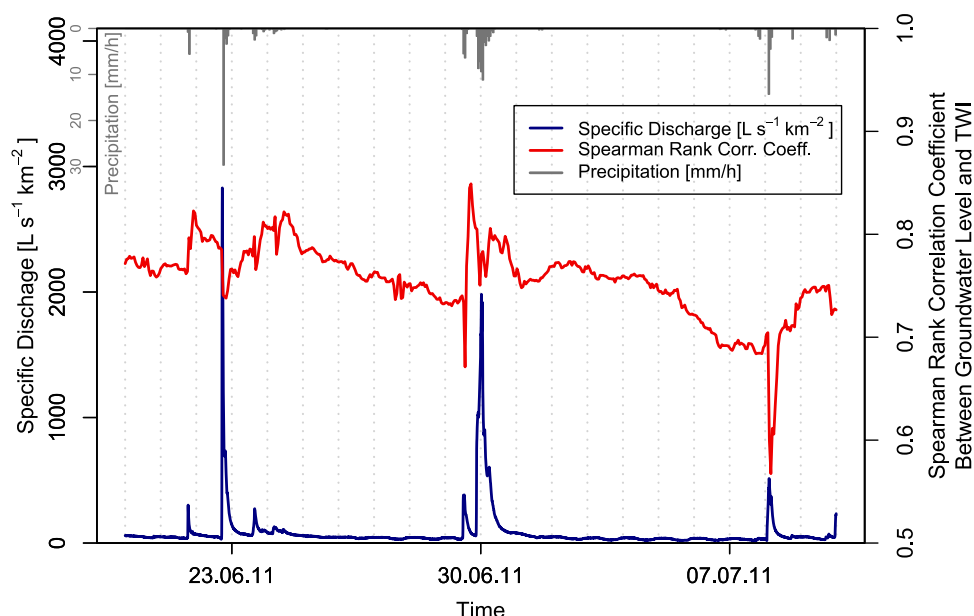


**Figure 4.** Median groundwater levels relative to soil depth (1 = at the soil surface, 0 = at bottom of the well) as a function of (a) local slope, (b) mean slope of the upslope contributing area, (c) local curvature, (d) mean curvature of the upslope contributing area, (e) upslope contributing area, and (f) Topographic Wetness Index.

The median relative groundwater levels increased linearly with TWI ( $r_s = 0.78$ ) but were highly variable for sites with a TWI between 4 and 6 (Figure 4f and Table 1). Sites with a TWI  $> 6$  had a median relative groundwater level of 0.7 or higher. The median relative groundwater levels were also correlated to the mean TWI of the upslope contributing area, but the correlation coefficient was lower ( $r_s = 0.62$ ). The correlations were similar for the median absolute groundwater levels (Table 1).

We also considered the soil depth and the saturated hydraulic conductivity of the mineral soil to be important controls on median groundwater levels, but the correlations were not statistically significant. The spatial distribution of soil type and vegetation within the study catchment was related to the median groundwater levels ( $p$  value of Mann-Whitney test  $< 0.001$ ) and could be predicted by topographic position, e.g., footslopes or depressions had predominantly mollic Gleysols and grassland vegetation, whereas ridge





**Figure 5.** Example of a time series of the Spearman rank correlation coefficients ( $r_s$ ) between groundwater levels at 51 locations and the Topographic Wetness Index (TWI) (red line). Precipitation and specific discharge at subcatchment C5 are shown in gray and blue, respectively.

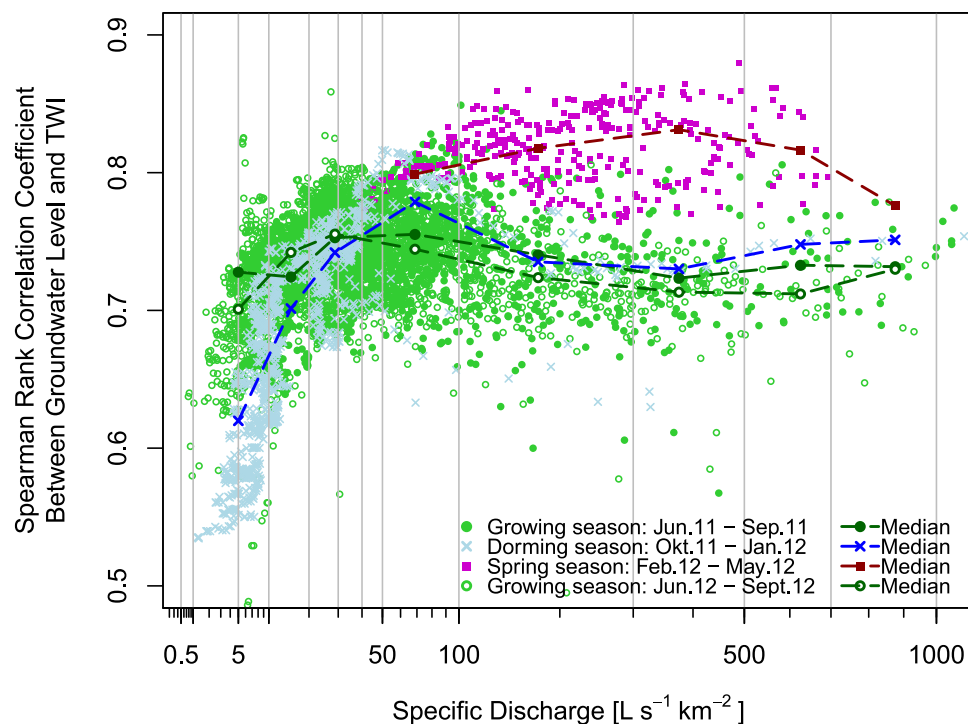
sites had predominantly umbric Gleysols and were often forested (Pearson's chi-square test,  $p < 0.001$  (soil type), and  $p < 0.003$  (vegetation); Cramer's V value, a measure of the strength of correlation, was 0.51 (soil type) and 0.33 (vegetation)).

### 3.3. Changes in the Correlation Between Groundwater Level Patterns and TWI Over Time

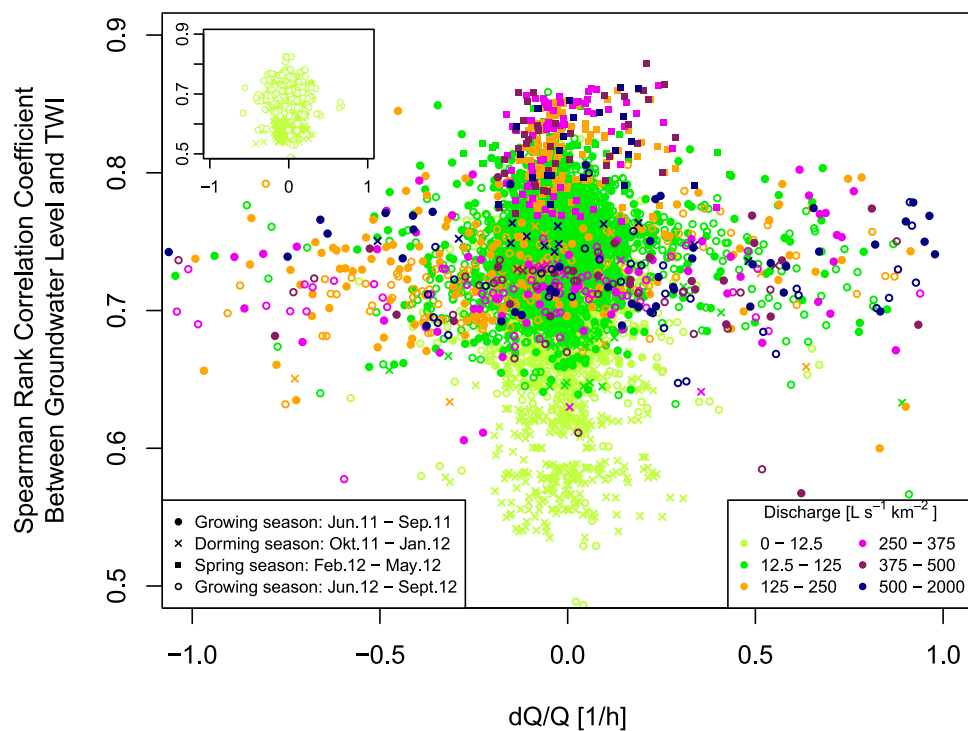
The continuous groundwater measurements allowed quantification of the temporal variation in the correlation between groundwater levels and topographic indices. The correlation between TWI and absolute groundwater levels decreased strongly at the beginning of rainfall events and reached the lowest values shortly after peak streamflow (Figure 5). During the falling limb of the hydrograph,  $r_s$  increased quickly and reached the highest values 12 h–2 days after the event. During dry periods,  $r_s$  gradually decreased until the beginning of the next event. The drop in correlation at the beginning of a rainfall event was particularly large after long, dry periods.

This event-scale change in correlation persisted throughout the year but was superimposed on a seasonal cycle (Figure 6):  $r_s$  was highest during spring, with values ranging between 0.75 and 0.85. Streamflow was never below  $40 \text{ L s}^{-1} \text{ km}^{-2}$  during spring. The lowest  $r_s$  of 0.5–0.6 occurred during the dormant season, in particular when streamflow was below  $10 \text{ L s}^{-1} \text{ km}^{-2}$ . This streamflow was exceeded during 87% of time during the 27 month study period. During the growing season, the correlation between groundwater levels and TWI varied between 0.65 and 0.75 during low ( $< 10 \text{ L s}^{-1} \text{ km}^{-2}$ ) and high ( $> 100 \text{ L s}^{-1} \text{ km}^{-2}$ ) streamflow conditions. These discharge values were exceeded during 87% and 13% of the study period, respectively. The maximum  $r_s$  of up to 0.80 occurred during intermediate streamflow conditions ( $10\text{--}100 \text{ L s}^{-1} \text{ km}^{-2}$ ; median streamflow:  $28 \text{ L s}^{-1} \text{ km}^{-2}$ ).

The wide range of streamflow conditions for which  $r_s$  values were higher than 0.7 suggested that it was not the event magnitude but rather conditions with small changes in runoff and, thus, also groundwater levels for which  $r_s$  values were highest. Under these conditions, the assumption of groundwater levels following a succession of steady state situations might have been fulfilled best. A bell-shaped relation with highest  $r_s$  values at near-zero  $dQ/Q$  (Figure 7) was pronounced for all streamflow conditions although for the smallest streamflow class ( $< 12.5 \text{ L s}^{-1} \text{ km}^{-2}$ ) it was least pronounced (see Figure 7, inset). Most of the low Spearman rank correlation coefficients in this class occurred during the dormant season, which is in agreement with the results shown in Figure 6.



**Figure 6.** Spearman rank correlation coefficients ( $r_s$ ) between groundwater level and Topographic Wetness Index (TWI) plotted as a function of specific discharge as an indicator of the average catchment state. Discharge from subcatchment C5 was chosen because it has the longest data series. The different colors and symbols indicate the different seasons. The median curves for defined streamflow classes are shown in darker dashed lines.



**Figure 7.** Spearman rank correlation coefficients ( $r_s$ ) between groundwater level and Topographic Wetness Index (TWI) plotted as a function of the relative change in streamflow at subcatchment C5. Symbols represent different seasons, while colors represent different streamflow classes. The inset in the upper left corner shows the data for streamflow  $< 12.5 \text{ L s}^{-1} \text{ km}^{-2}$  without overlap of the other streamflow classes.

## 4. Discussion

### 4.1. The Role of Topography on Groundwater Levels

The statistical significance and strength of the correlation ( $r_s > 0.6$ ) suggest that topography exerts a significant control on the median groundwater levels in mountain catchments with low-permeability soils. Median groundwater levels were related to local controls, such as the local slope and the soil wetness (as described by the TWI), and upslope controls, such as the runoff concentration within the upslope contributing area (as described by the mean upslope curvature), subsurface water input from upslope (as described by the upslope contributing area), and mean soil wetness in the source area (as described by the mean TWI of the upslope contributing area). Interestingly, the relative strength of slope and curvature in explaining the median groundwater levels depended on whether they were considered as local or upslope controls.

Other studies also reported groundwater levels to be correlated to TWI, although the correlation coefficients were lower than in our study [Detty and McGuire, 2010]. A possible explanation for the lower correlations might be the more permeable soils in these catchments, which might lead to deeper median groundwater levels that are less influenced by the surface topography [Bachmair and Weiler, 2012]. In other studies, particularly on footslopes and in catchments with a relatively flat topography or conductive soils, topography was not identified as a dominant control and the TWI was weakly correlated to spatial groundwater level variations [Moore and Thompson, 1996; Seibert et al., 1997]. This is plausible since in flatter sites the hydraulic gradient, subsurface flow concentration, and contribution from upslope areas are smaller and, therefore, other controls are more likely to dominate the variability in median groundwater levels.

Only a few other studies have commented on the correlation between groundwater levels and topographic controls other than TWI. Bachmair and Weiler [2012] reported a nonsignificant correlation between local plan- and profile curvature and mean relative groundwater levels but did not investigate curvature of the upslope contributing area. The local slope was among the predictor variables with the strongest correlation ( $r_s = -0.36$ ) with mean relative groundwater levels, but the correlation was lower than in our study ( $r_s = -0.69$ ). Other predictor variables with similar or slightly higher correlations were land use ( $r_s = -0.42$ ) and saturated hydraulic conductivity ( $r_s = -0.39$ ). Bachmair and Weiler [2012] concluded, based on the low correlation coefficients, that important predictor variables were missing in their analysis but that topography and soil properties were among the important controls on groundwater responses of the three experimental hillslopes with transmissive soils. This is noticeable as their experimental hillslopes were explicitly chosen to be relative planar. Soil depth and saturated hydraulic conductivity of the mineral soil were not correlated to median groundwater levels in this hillslope study.

The upslope contributing area exerts an important control on groundwater and subsurface flow. Previous studies in the Alptal concluded that these fluxes were important components of the hillslopes water balance [Feyen et al., 1996]. Subsurface runoff (364 mm) from a small 10 m<sup>2</sup> experimental plot with 80 cm deep PVC panels on the uphill side and a trench on the downhill side exceeded net precipitation (=128 mm precipitation minus 26 mm evapotranspiration) by more than 260 mm during an 11 day measurement campaign [Feyen et al., 1996]. While this example might be exceptional due to groundwater upwelling at that topographic location, it shows that subsurface water input from upslope areas can be substantial. Bachmair and Weiler [2012] reported upslope contributing area to be more important than vegetation and soil properties only when accounting for interactions between predictor variables but not in the partial correlation analysis. Detty and McGuire [2010] found upslope contributing area to be significantly related to catchment wide water table duration but not when the analysis was performed for individual landforms (footslope, midslope, shoulder) or well transects.

The fact that other studies reported a lower correlation between topography and groundwater levels suggests that the governing subsurface runoff processes may be different in contrasting catchments. In steep mountain headwater catchments with low-permeability soils (e.g., Gleysols), perched groundwater systems are expected to prevail. As groundwater levels are predominantly shallow, subsurface flow through conductive soil layers and/or preferential flow paths near the soil surface is likely an important flow component during events. The humid conditions, together with the low drainable porosity of the soil matrix, cause median groundwater levels to be persistently close to the soil surface, soil moisture to be high, and storage capacity to be low. Our results and field observations suggest that spatial variability in groundwater levels is driven by the input from upslope areas, which is influenced by subsurface flow concentration (convergent or divergent shallow flow pathways in

the upslope contributing area of each site). The local hydraulic gradient exerts a control on the downslope drainage conditions, which, together with upslope soil water inputs, determine groundwater levels.

In terms of differences in the dominant controls on groundwater levels and runoff mechanisms in different catchments, it appears that saturation and subsequent lateral subsurface flow in transmissive soils occurs at deeper depth than in low-permeability soils. Therefore, soil properties, like the saturated hydraulic conductivity and soil depth, as well as topography and infiltrability of the bedrock surface or deep impeding soil layer, are expected to be of greater importance than in environments with low-permeability soils [McDonnell, 1990; Uchida *et al.*, 2003; Tromp-van Meerveld *et al.*, 2007].

#### 4.2. Predictability of Median Groundwater Levels

Variability in median groundwater levels was largest for sites with a local slope between 30 and 50% (24 sites out of 51), an upslope contributing area between 200 and 600 m<sup>2</sup> (18 sites out of 51), and a TWI between 4 and 6 (27 sites out of 51). These criteria applied to relatively large parts of the catchment (49%, 32%, 49%, respectively), predominantly at midslope locations. Eleven out of the 51 sites fulfilled all three criteria; two of them were among the most responsive sites in the catchment with the largest groundwater amplitude. The median groundwater levels were not statistically significantly different for the umbric and mollic Gleysols (Mann-Whitney test,  $p > 0.28$ ), which suggests that soil type did not cause the large variability in median groundwater levels in this zone. Flatter footslopes and steeper ridge sites were characterized by a smaller variability in median groundwater levels. It could be speculated that a more complex and, therefore, more variable interplay of several, well-correlated controls dominate median groundwater levels on the midslopes, while for the footslopes and ridges only a few important factors determine the balance between subsurface input from upslope and drainage. This makes prediction of median groundwater levels in footslope and ridge sites more reliable than for midslopes and suggests that midslopes are most relevant in terms of monitoring changes in groundwater storage and hydrological connectivity.

#### 4.3. TWI Assumptions Evaluated by the Temporal Variability of Correlation Strength

The spatial groundwater level pattern did not maintain a persistent shape that shifted uniformly up and down in response to changes in saturated zone storage as assumed by the physical motivation of using TWI for modeling groundwater levels or streamflow. Instead the spatial pattern in groundwater levels changed during events and seasonally. We hypothesize that temporal differences in rainfall inputs and spatio-temporal differences in soil water storage cause differences in groundwater responses throughout the catchment during a rainfall event. While we expect parts of the catchment to be hydrologically disconnected prior to events or during dry periods, we assume large parts of the upslope contributing area to be connected during events (see TWI assumptions). In these situations, water tables are high and the local slope is a good predictor of the hydraulic gradient. During recession, groundwater levels slowly decline and the assumption of a succession of steady state conditions is more realistic, which was also indicated by stronger correlations during these periods. Toward the end of the recession period, parts of the upslope contributing area might become hydrologically disconnected. The longer the time that groundwater levels fall, the more heterogeneous they become throughout the catchment and the weaker the correlation with TWI becomes. Sites with a large upslope contributing area or low slope tend to have persistently high groundwater levels, while wells at other sites can fall dry.

The assumption of a persistent shape of the groundwater pattern that shifts uniformly up and down due to changes in saturated zone storage did hold neither for events nor for seasons. During the growing season, the groundwater pattern within the catchment varied because it was determined by differences in groundwater response during rainfall events. During the longer dry periods in late fall and winter, differences in groundwater levels were most pronounced. The TWI assumptions could be considered to be reasonably met only toward the end of the snowmelt season, when constant, low-intensity melt water inputs throughout large parts of the catchment caused groundwater levels to be high, and the upslope contributing area was, therefore, likely to be hydrologically connected (Figures 6 and 7).

More generally speaking, the assumption of steady state successions was best met during conditions of small changes in runoff (= near-zero  $dQ/Q$ ) and presumably small changes in groundwater levels (Figure 7). The assumptions were, however, not fulfilled during large changes in groundwater levels and streamflow

during the start of events, when spatial variability of rainfall inputs and subsurface flow from upslope areas, drainage, and associated delays were high. The saturated zone did also not respond in unison during the lowest flows at the end of long dry periods, when some wells were dry and connectivity was likely lowest. This was particularly pronounced during the long dry period in winter (see Figure 6, light blue data points, and Figure 7, light green data points).

## 5. Concluding Remarks

We found that topography is a good predictor of median groundwater levels in the studied mountain headwater catchment with low-permeability soils. Median groundwater levels were correlated with topographic indices and the strength of correlation differed depending on whether they were considered a local or an upslope topographic control. This suggests that groundwater levels were not only controlled by local drainage but also by subsurface inputs from upslope and that both scales (local and upslope contributing area) have to be considered to better understand the spatial variability in median groundwater levels.

This study also showed that the rank correlation between groundwater levels and TWI was not constant over time but decreased during rainfall events as differences in rainfall input and subsurface flow redistribution and associated delays led to spatial differences in groundwater responses. When groundwater levels were high and changed slowly, e.g., when the catchment was slowly draining after events or during snowmelt in spring, the TWI assumptions of steady state successions, connected upslope contributing areas, and surface slope as a proxy of the hydraulic gradient were fulfilled best. They were least appropriate during long dry periods, when parts of the catchment drained differently and became disconnected.

We expect our findings to also be applicable in other humid mountain headwater catchments with low-permeability soils and shallow groundwater tables as the topographic indices are proxies for generally applicable, physical properties and processes that seem to dominate in these catchments. Our study showed that the TWI assumptions might be useful simplifications for modeling applications in catchments with shallow groundwater levels during periods following rainfall events and during the snowmelt season when streamflow and groundwater levels change slowly. However, for modeling the groundwater response at the beginning of events and during long dry periods other modeling approaches are needed to better represent the saturated zone dynamics. This has implications for using TWI-based models to predict the spatial patterns of groundwater levels, their connectivity, and catchment runoff response.

## Acknowledgments

The authors thank Manfred Stähli for helpful discussions and supply of meteorological data, as well as Russell Smith and Dan Moore for valuable thoughts on the measurement concept and data analysis. Special thanks to Benjamin Fischer for sharing data, support and encouragement in the field, and numerous fruitful discussions. We thank the editor and the three reviewers for constructive comments on the manuscript, Tracy Ewen for proofreading Ellen Cerwinka, Jana Dusik, Seraina Kauer, Nadja Lavanga, Claudia Müller, Stephan Müller, Stefan Plötner, Sandra Pool, Paribesh Pradhan, Sandra Schäfer, Karl Steiner, and Ivan Woodrich for field and laboratory assistance, Nans Addor for R tips, Volker Wichmann and Ronald Schmidt for SAGA and GIS data support, and the Swiss Federal Institute for Forest, Snow and Landscape Research WSL as well as OAK - Gemeinde Alpthal (canton Schwyz) for cooperation.

## References

- Ambrose, B. (2004), Variable "active" versus "contributing" areas or periods: A necessary distinction, *Hydrol. Processes*, 18(6), 1149–1155, doi:10.1002/hyp.5536.
- Anderson, M. G., and T. P. Burt (1978), The role of topography in controlling throughflow generation, *Earth Surf. Processes*, 3, 331–334.
- Bachmair, S., and M. Weiler (2012), Hillslope characteristics as controls of subsurface flow variability, *Hydrol. Earth Syst. Sci.*, 16(10), 3699–3715, doi:10.5194/hess-16-3699-2012.
- Barling, R. D., W. Corporation, I. D. Moore, and B. Grayson (1994), A quasi dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content, *Water Resour. Res.*, 30(4), 1029–1044.
- Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24(1), 43–69.
- Bouwer, H., and R. C. Rice (1976), A test for determining hydraulic conductivity of unconfined aquifers with completely or partially penetrating wells, *Water Resour. Res.*, 12(3), 423–428.
- Breiman, L. E. O. (2001), Random forests, *Mach. Learning*, 45, 5–32.
- Burt, T. P., and D. P. Butcher (1985), Topographic controls of soil moisture distributions, *J. Soil Sci.*, 36(1978), 469–486.
- Buttle, J. M., P. W. Hazlett, C. D. Murray, I. F. Creed, D. S. Jeffries, and R. Semkin (2001), Prediction of groundwater characteristics in forested and harvested basins during spring snowmelt using a topographic index, *Hydrol. Processes*, 15(18), 3389–3407, doi:10.1002/hyp.1033.
- Conrad, O. (2007), Entwurf, Funktionsumfang, und Anwendung eines Systems für Automatisierte Geowissenschaftliche Analysen, PhD dissertation, 221 pp., Univ. of Göttingen, Göttingen, Germany.
- Desbarats, A. J., C. E. Logan, M. J. Hinton, and D. R. Sharpe (2002), On the kriging of water table elevations using collateral information from a digital elevation model, *J. Hydrol.*, 255(1–4), 25–38, doi:10.1016/S0022-1694(01)00504-2.
- Detty, J. M., and K. J. McGuire (2010), Topographic controls on shallow groundwater dynamics: Implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment, *Hydrol. Processes*, 24(16), 2222–2236, doi:10.1002/hyp.7656.
- Evans, S. (1980), An integrated system of terrain analysis and slope mapping, *Z. Geomorphol.*, supplement, 36, 274–294.
- Feyen, H., J. Leuenberger, A. Papritz, M. Gysi, H. I. Ilihl, and P. Schleppi (1996), Runoff processes in catchments with a small scale topography, *Phys. Chem. Earth*, 21(3), 177–181.
- Feyen, H., H. Wunderli, H. Wydler, and A. Papritz (1999), A tracer experiment to study flow paths of water in a forest soil, *J. Hydrol.*, 225(3–4), 155–167.
- Gomi, T., R. C. Sidle, S. Miyata, K. Kosugi, and Y. Onda (2008), Dynamic runoff connectivity of overland flow on steep forested hillslopes: Scale effects and runoff transfer, *Water Resour. Res.*, 44, W08411, doi:10.1029/2007WR005894.

- Hagedorn, F., P. Schleppi, P. Waldner, and H. Flüeler (2000), Export of dissolved organic carbon and nitrogen from Gleysol dominated catchments—The significance of water flow paths, *Biogeochemistry*, 50, 137–161.
- Hagedorn, F., P. Schleppi, J. Bucher, and H. Flüeler (2001), Retention and leaching of elevated N deposition in a forest ecosystem with Gleysols, *Water Air Soil Pollut.*, 129, 119–142.
- Haught, D. R. W., and H. J. van Meerveld (2011), Spatial variation in transient water table responses: Differences between an upper and lower hillslope zone, *Hydrol. Processes*, 25(25), 3866–3877, doi:10.1002/hyp.8354.
- Hegg, C., B. W. McArdeall, and A. Badoux (2006), One hundred years of mountain hydrology in Switzerland by the WSL, *Hydrol. Processes*, 20(2), 371–376, doi:10.1002/hyp.6055.
- Hewlett, J., and A. Hibbert (1967), Factors affecting the response of small watersheds to precipitation in humid areas, in *Proceedings of First International Symposium on Forest Hydrology*, vol. 33, edited by W. E. Sopper and H. W. Lull, pp. 275–290, Pergamon, Oxford.
- Laudon, H., V. Sjöblom, I. Buffam, J. Seibert, and M. Morth (2007), The role of catchment scale and landscape characteristics for runoff generation of boreal streams, *J. Hydrol.*, 344(3–4), 198–209, doi:10.1016/j.jhydrol.2007.07.010.
- Lehmann, P., C. Hinz, G. McGrath, H. J. Tromp-van Meerveld, and J. J. McDonnell (2007), Rainfall threshold for hillslope outflow: An emergent property of flow pathway connectivity, *Hydrol. Earth Syst. Sci.*, 11(2), 1047–1063.
- McDonnell, J. J. (1990), A rationale for old water discharge through macropores in a steep, humid catchment, *Water Resour. Res.*, 26(11), 2821–2832.
- McGlynn, B. L., J. J. McDonnell, J. Seibert, and C. Kendall (2004), Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations, *Water Resour. Res.*, 40, W07504, doi:10.1029/2003WR002494.
- Mohn, J., A. Schürmann, and F. Hagedorn (2000), Increased rates of denitrification in nitrogen-treated forest soils, *For. Ecol. Manage.*, 137, 113–119.
- Moore, R. D., and J. C. Thompson (1996), Are water table variations in a shallow forest soil consistent with the TOPMODEL concept?, *Water Resour. Res.*, 32(3), 663–669, doi:10.1029/95WR03487.
- O'Loughlin, E. M. (1986), Prediction of surface saturation zones in natural catchments by topographic analysis, *Water Resour. Res.*, 22(5), 794–804.
- Penna, D., N. Mantese, L. Hopp, M. Borga, and G. Dalla Fontana (2014), Spatio-temporal variability of piezometric response on two steep alpine hillslopes, *Hydrol. Processes*, doi:10.1002/hyp, in press.
- R Development Core Team (2005), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna, Australia.
- Rodhe, A., and J. Seibert (2011), Groundwater dynamics in a till hillslope: Flow directions, gradients and delay, *Hydrol. Processes*, 25(12), 1899–1909, doi:10.1002/hyp.7946.
- Schleppi, P., N. Muller, H. Feyen, A. Papritz, J. B. Bucher, and H. Flüeler (1998), Nitrogen budgets of two small experimental forested catchments at Alptal, Switzerland, *For. Ecol. Manage.*, 101, 177–185.
- Seibert, J., and B. L. McGlynn (2007), A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models, *Water Resour. Res.*, 43, W04501, doi:10.1029/2006WR005128.
- Seibert, J., K. Bishop, and L. Nyberg (1997), A test of TOPMODEL's ability to predict spatially distributed groundwater levels, *Hydrol. Processes*, 11, 1131–1144.
- Seibert, J., K. Bishop, A. Rodhe, and J. J. McDonnell (2003), Groundwater dynamics along a hillslope: A test of the steady state hypothesis, *Water Resour. Res.*, 39(1), 1014, doi:10.1029/2002WR001404.
- Spearman, C. (1904), The proof and measurement of association between two things, *J. Psychol.*, 15, 72–101, doi:10.1093/ije/dyq191.
- Spence, C., and M. Woo (2006), Hydrology of subarctic Canadian Shield: Heterogeneous headwater basins, *J. Hydrol.*, 317(1–2), 138–154, doi:10.1016/j.jhydrol.2005.05.014.
- Spence, C., X. J. Guan, R. Phillips, N. Hedstrom, R. Granger, and B. Reid (2009), Storage dynamics and streamflow in a catchment with a variable contributing area, *Hydrol. Processes*, 24(16), 2209–2221, doi:10.1002/hyp.7492.
- Tarboton D. G. (1997), A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resour. Res.*, 33–2, pp. 309–319.
- Travis, M. R., W. D. Iverson, H. Gary, and C. G. Johnson (1975), VIEWIT: Computation of seenarea, slope and aspect for land-use planning PSW 11, *Gen. Tech. Rep. PSW-GTR-11*, U.S. Dep. Agric. – For. Serv., Berkeley, Calif.
- Troch, P. A., M. Mancini, C. Paniconi, and E. F. Wood (1993), Evaluation of a distributed catchment scale water balance model, *Water Resour. Res.*, 29(6), 1805–1817.
- Tromp-van Meerveld, H. J., and J. J. McDonnell (2006), Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, *Water Resour. Res.*, 42, W02411, doi:10.1029/2004WR003800.
- Tromp-van Meerveld, H. J., N. E. Peters, and J. J. McDonnell (2007), Effect of bedrock permeability on subsurface stormflow and the water balance of a trenched hillslope at the Panola Mountain Research Watershed, Georgia, USA, *Hydrol. Processes*, 21, 750–769, doi:10.1002/hyp.
- Uchida, T., Y. Asano, N. Ohte, and T. Mizuyama (2003), Seepage area and rate of bedrock groundwater discharge at a granitic unchanneled hillslope, *Water Resour. Res.*, 39(1), 1018, doi:10.1029/2002WR001298.
- U.S. Department of the Interior (2001), *Water Measurement Manual*, 3rd ed., U.S. Gov. Print. Off., Washington, D. C.
- Zehe, E., R. Becker, A. Bardossy, and E. Plate (2005), Uncertainty of simulated catchment runoff response in the presence of threshold processes: Role of initial soil moisture and precipitation, *J. Hydrol.*, 315(1–4), 183–202, doi:10.1016/j.jhydrol.2005.03.038.

# PAPER V





# Groundwater response timing in a pre-alpine catchment

M. Rinderer<sup>1\*</sup>, H. J. van Meerveld<sup>1</sup>, M. Stähli<sup>2</sup>, J. Seibert<sup>1,3</sup>

<sup>1</sup> Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland; <sup>2</sup> Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland; <sup>3</sup> Department of Earth Sciences, Uppsala University, SE-752 36 Uppsala, Sweden

\*Corresponding Author: [michael.rinderer@geo.uzh.ch](mailto:michael.rinderer@geo.uzh.ch)

---

## Abstract

Groundwater levels in steep headwater catchments typically respond quickly to rainfall but the response might vary spatially across the catchment. In this study we investigated the topographic controls and the effect of rainfall and antecedent conditions on groundwater response timing for 133 rainfall events in a 20 ha pre-alpine catchment with low permeability soils in Switzerland. The median time to rise and median duration of recession of the 51 groundwater monitoring sites were highly correlated to the topographic characteristics of the site and its upslope contributing area. The median time to rise depended more on the topographic characteristics than on the rainfall characteristics or antecedent soil wetness conditions. The median time to rise decreased with Topographic Wetness Index (TWI) for sites with a TWI < 6 and was almost constant for sites with higher TWI. The slope of this relation was a function of rainfall intensity. The rainfall threshold for groundwater initiation was also a function of TWI and allowed extrapolation of point measurements to the catchment scale. The median lag time between the centroid of rainfall and the groundwater peak was 75 minutes. Half of the groundwater levels peaked before the streamflow peak at the catchment outlet but only by 15 to 25 minutes. The stronger correlations between topographic indices and groundwater response timing characteristics in this study compared to previous studies suggest that surface topography affects the groundwater response timing in catchments with low permeability soils more than in catchments with more transmissive soils.

Keywords: groundwater, response timing, TWI, topographic controls, subalpine catchment, rainfall threshold, antecedent wetness, spatial patterns

---

## 1. Introduction

In steep mountain headwater catchments, shallow groundwater can respond quickly to rainfall because alpine soils are typically thin and gradients are steep. These groundwater dynamics play an important role in runoff generation and hydrologic connectivity of the hillslopes to the stream because they exert a strong control on lateral subsurface stormflow (Weiler et al. 2005). Identifying the factors that control the spatial variability in shallow groundwater dynamics will, therefore, improve our understanding of how catchments function (McGlynn & McDonnell 2003; McDonnell et al. 2007).

Because the magnitude and timing of the groundwater response to rainfall are controlled by different variables, such as topography and soil- and bedrock properties, the occurrence of perched groundwater can be very patchy and the response to rainfall can be highly variable (Penna et al. 2014; Bachmair & Weiler 2012; Fannin et al. 2000). Previous measurements of groundwater levels across hillslope transects and catchments have revealed that the timing and magnitude of the water table response

is related to landform (Detty & McGuire 2010), distance to the stream channel network (Seibert et al. 2003; Rodhe & Seibert 2011; Haught & van Meerveld 2011), thickness of the soil or the topography of the bedrock (Penna et al. 2014; Tromp-van Meerveld & McDonnell 2006).

The observations reported in the literature are, however, ambiguous with respect to the correlation between groundwater levels and streamflow. In some catchments, groundwater levels close to the stream were well correlated with each other and with discharge, but groundwater levels in upslope locations were not (Haught & van Meerveld 2011; Seibert et al. 2003). The decreasing correlation between groundwater levels and discharge with increasing distance from the stream suggested that upslope areas did not contribute directly to streamflow during events. Furthermore, sites close to the stream responded prior to streamflow, while the groundwater response in the upslope sites was delayed and more variable. As antecedent soil water content increased, groundwater lag times became shorter and groundwater peaks preceded streamflow peaks (Haught & van Meerveld 2011). Other studies have also shown that the runoff response precede the

groundwater response (Penna et al. 2014), which at a first glance is contradictory to the common perception of how groundwater contributes to streamflow (Sklash & Farvolden 1979). Yet other researchers have shown that groundwater response times were shortest in the upper parts of the hillslopes and catchments and related this to the spatial distribution of soil thickness and the topography of the soil-bedrock interface (McDonnell 1990; Rodhe & Seibert 2011; Penna et al. 2014; Tromp-van Meerveld & McDonnell 2006). But for other catchments there was no correlation between the duration of transient saturation and the distance from the stream (Lana-Renault et al. 2013) and no relation between peak groundwater level and topographic position (Dhakal & Sullivan 2014). However, the instrumentation in some studies was limited to the interface between the hillslope and the riparian zone and results may therefore not be representative for catchment-wide groundwater dynamics (Anderson & Burt 1978; Moore & Thompson 1996).

These partly contradictory observations reflect site-specific settings and have made it difficult to generalize these findings or to transfer them to other catchments. Nevertheless, attempts were made to explain groundwater responses based on catchment characteristics such as topography, soil properties or vegetation. Under wet environmental conditions (Anderson & Burt 1978; Burt & Butcher 1985; Lana-Renault et al. 2013), steep terrain (Penna et al. 2014) or shallow groundwater tables (Troch et al. 1993), variability in groundwater responses were related to topography. Under dry conditions (Detty & McGuire 2010), flat terrain (Barling et al. 1994) and especially in permeable soils (Seibert et al. 1997; Dhakal & Sullivan 2014; Anderson et al. 2010), the relation between groundwater response and topography was not clear. In catchments with transmissive soils, the variability in saturated hydraulic conductivity, soil depth, bedrock topography, vegetation distribution and snowmelt patterns could explain the variability in groundwater response better than topography (Bachmair & Weiler 2012; Smith et al. 2013).

Rainfall input and antecedent conditions are also important controls on shallow groundwater responses. Groundwater peak duration and response amplitude were larger during the wet season and during events that exceeded a certain rainfall threshold in the Hubbard Brook Experimental Forest catchment in the New Hampshire, USA (Detty & McGuire 2010). On the contrary, in the Black Forest in Germany groundwater responses were small and slow during wet conditions in fall, winter and spring and affected predominantly the footslopes, while

during dry summer conditions the groundwater responses were quicker, more variable and occurred across the whole hillslope (Bachmair et al. 2012). For other hillslopes or catchments, the percentage of groundwater wells that showed a response during individual rainfall events was correlated to total event precipitation and storm duration but not to rainfall intensity and antecedent conditions (Penna et al. 2014; Dhakal & Sullivan 2014; Fannin et al. 2000).

Despite the knowledge gained by these hillslope-scale studies at sites with transmissive soils, we still know little about catchment-scale groundwater dynamics in steep mountain environments with less permeable soils. One might expect the groundwater levels to be closer to the surface and to be more responsive to rainfall because of the lower storage deficit, low drainable porosity and low hydraulic conductivity of the mineral soil. As groundwater levels rise close to the soil surface and into higher permeability soil layers, surface topography might exert a stronger control on the lateral redistribution of water (Hutchinson & Moore 2000). One could therefore expect surface topography to explain a larger fraction of the variability in shallow groundwater responses in a catchment with low permeability soils than has been shown in previous studies for catchments with higher permeability soils. To test this assumption, we analyzed the timing of the groundwater responses in a subalpine headwater catchment in Switzerland and correlated it to topographic indices and rainfall and antecedent wetness conditions.

In particular, we address the following questions:

- (i) To what extent does topography govern the timing of the groundwater response, in particular the start of the groundwater level rise, the timing of peak groundwater level and duration of the recession?
- (ii) Is there a rainfall threshold for groundwater response initiation and if so, does this threshold depend on the topography?
- (iii) How do antecedent soil wetness conditions and rainfall intensity influence the timing of the groundwater response?

## 2. Methods

### 2.1 Study Catchment

The 20 ha study catchment is located in the Alptal, a pre-alpine valley about 40 km southeast of Zurich, Switzerland (Fig. 1). The catchment is steep with an average slope of 35 % and extends from 1270 m asl. to 1650 m asl. Mean annual precipitation in the

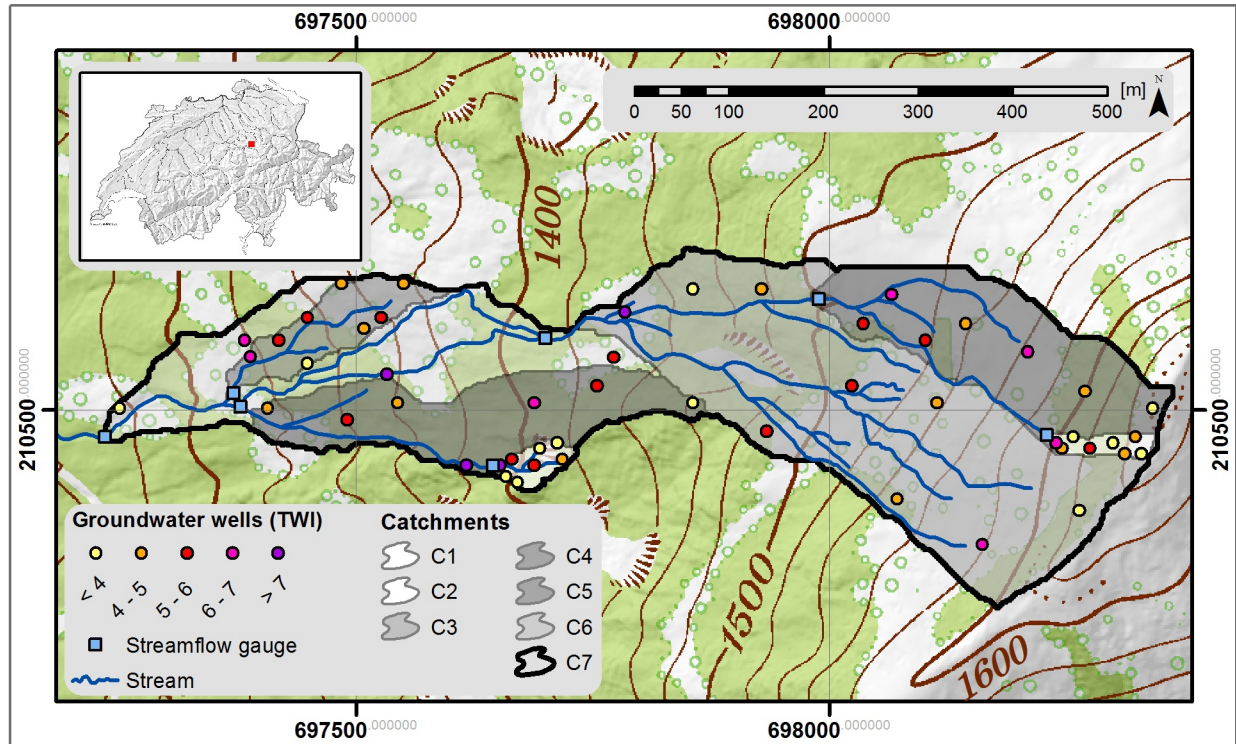


Fig. 1: Map of the experimental catchment showing the seven nested sub-catchments with a streamflow gauging station at each outlet and the location of the 51 spatially distributed groundwater wells. Groundwater wells are color-coded according to the Topographic Wetness Index (TWI). (Background-topographic map: Swisstopo, 123456789).

region is 2300 mm/year, and about 30 % falls as snow (Feyen et al. 1999). The catchment is normally snow-covered between December and May and the largest and most intense rainfall events occur typically between June and September. The catchment is characterized by a distinct small-scale topography with hollows and ridges and a dense natural drainage network (205 m/ha). The main channel close to the catchment outlet has 2 to 4 m deep banks on both sides but the other streams are not deeply incised. A distinct riparian zone is missing in this study catchment. The Topographic Wetness Index (TWI) (Beven & Kirkby 1979) varies between 2 and 14 (median TWI: 5): 19 % of the catchment has a TWI < 4; 49 % of the catchment has a TWI between 4 and 6; and 32 % of the catchment has a TWI > 6. Moor landscapes and wet grassland areas are common in hollows and flatter parts of the catchment (ca. 7 ha), while on steeper slopes and ridge-sites open coniferous forest grow (*Picea abies* L. with an understory of *Vaccinium* sp.; ca. 11 ha) (Hagedorn et al. 2000). Parts of the upper catchment (ca. 2 ha) is seasonally used for grazing cattle. In wet depressions where the water table is persistently close to the soil surface, the soils are mollic Gleysols with

a topsoil high in carbonate. The mineral soil consists of a permanently reduced Bg horizon, with typically 43 % clay, 42 % silt and 15 % sand (Schleppi et al. 1998). At the ridge sites, where the water table is normally more than 0.40 m below the soil surface, the soils are umbric Gleysols with an oxidized Bw horizon (49 % clay, 46 % silt and 5 % sand) (Schleppi et al. 1998; Hagedorn et al. 2001). Soil depth varies between 0.5 m at ridge sites to more than 2.5 m in depressions. The bedrock consists of a poorly permeable clay-rich Flysch with calcareous sandstone and argillite and bentonite schist layers (Mohn et al. 2000).

## 2.2 Field Measurements

Groundwater levels were measured continuously at 51 locations across the study site between September 2010 and the end of November 2012. The monitoring sites were selected based on a stratified random sampling approach using the TWI in seven nested sub-catchments (C1 to C7, ranging in size from ~0.2 ha to 20 ha; Fig. 1). This procedure guaranteed representative sampling of the range of topographic positions, soil types and vegetation in the

experimental catchment (8 ridge site, 22 midslope- and 21 footslope- or depression locations; 25 mollic Gleysol sites and 26 umbric Gleysol sites; 20 forested sites and 31 grassland sites). At each site, a borehole was manually drilled down to refusal (mean well depth: 1.06 m, min: 0.46 m, max: 2.16 m). The boreholes were fitted with a 4 cm diameter PVC pipe, screened over the full length up to 10 cm below the surface and backfilled with coarse filter sand. The filter pack was sealed with bentonite and plastic foil 5-10 cm below the surface to prevent water entering the well from the soil surface. Water levels were measured in the wells at a 5 min interval during summer (May to December) and a 10 min interval during winter using Odyssey capacitance water level loggers (Dataflow Systems Pty Limited). Groundwater level measurements were checked manually approximately every 2 to 3 months and corrected for a potential offset.

Stream stage at the outlet of the 20 ha study catchment (C7 in Fig. 1) was measured in a natural cross-section every 5 minutes from May to December 2011 and May and December 2012 using pressure loggers (DL/N 70 by STS, Sensor Technik Sirmach AG). Weir construction was not possible due to sediment transport but changes in the natural cross-section were documented monthly and deemed to be minor for the study period. Salt dilution measurements during seven events of different magnitude and a low flow period were used to determine the rating curve for the cross section. The rating curve covers 58 % of the range of water levels recorded during the study period and had to be extrapolated for only 1 % of the total study period.

The extrapolation is not considered to have a major impact on the results of this study as it mainly affects the size of the peakflows and not the timing of the response.

Precipitation and air temperature were recorded every 10 minutes and barometric pressure every 5 minutes at a permanent meteorological weather station 1 km from the experimental catchment at 1219 m asl. There was no reliable information on the spatial pattern of precipitation in the catchment but we expect the altitudinal gradient in precipitation to be small. For the correlation analysis, primarily median time lags were chosen in order to be less affected by potential errors due to spatial differences in the timing of the onset of rainfall for individual events.

### 2.3 Rainfall Event Characteristics

Rain events were defined as events exceeding 5 mm of total rainfall (the median daily rainfall of all days with rain) or had a maximum rainfall intensity > 2 mm/10min (the 85 % quantile), separated by at least 2 hours without rainfall. Events during winter (i.e., between December 1<sup>st</sup>, 2010 and April 12<sup>th</sup>, 2011 and between December 1<sup>st</sup>, 2011 and May 21<sup>st</sup>, 2012), when the catchment was snow-covered, were excluded from the analyses. The total rainfall during the 133 events that were analyzed was 3027 mm or 93 % of the total rainfall during the snow-free period of the two years considered (3262 mm). The selected rainfall events differed considerably in mean and maximum rainfall intensity, total amount of rainfall,

Tab. 1: Characteristics of the four rainfall event types: type 1a: low-intensity/dry, type 1b: low-intensity/wet, type 2a: moderate- intensity /dry, type 2b: moderate-intensity/wet. Similar superscript letters indicate which pairs are not significantly different based on a pairwise Mann-Whitney test and Bonferroni adjusted p-values.

	Rainfall Event Type				
	1a	1b	2a	2b	all
Intensity	low	low	moderate	moderate	-
Antecedent wetness	dry	wet	dry	wet	-
Number of events	30	27	33	43	133
Median average intensity [mm/h]	1.1 <sup>a</sup>	1.2 <sup>a</sup>	3.3 <sup>b</sup>	2.9 <sup>b</sup>	2.0
Median 3 day antecedent rainfall [mm]	2.1 <sup>a</sup>	27.9 <sup>b</sup>	1.7 <sup>a</sup>	22.4 <sup>b</sup>	11.1
Median maximum intensity [mm/10min]	1 <sup>a</sup>	1.4 <sup>a</sup>	2.5 <sup>b</sup>	2.8 <sup>b</sup>	1.8
Median event sum [mm]	11.3 <sup>a</sup>	14.5 <sup>ab</sup>	18 <sup>bc</sup>	20.6 <sup>ac</sup>	17.4
Median time to rainfall centroid [min]	320 <sup>a</sup>	360 <sup>a</sup>	130 <sup>b</sup>	210 <sup>ab</sup>	240
Median event duration [min]	670 <sup>a</sup>	720 <sup>a</sup>	330 <sup>b</sup>	480 <sup>ab</sup>	550

event duration and antecedent wetness conditions and were therefore subdivided into four rainfall event types: *Type 1a*: low-intensity/dry antecedent conditions, *Type 1b*: low-intensity/wet antecedent conditions, *Type 2a*: moderate-intensity/dry antecedent conditions, *Type 2b*: moderate-intensity/wet antecedent conditions. The class breaks were set at an event-average rainfall intensity of 1.8 mm/h and a 3 day sum of antecedent precipitation of 10 mm. These breaks reflect the mean event rainfall intensity that caused a water level response for at least 10 % of all sites (10 % quantile) and the median of the 3 day sum of antecedent precipitation. This classification resulted in roughly 30 events in each class (Tab. 1). Differences in event characteristics between the event classes were tested for statistical significance using the Mann-Whitney test with adjusted p-values based on the Bonferroni method.

The low-intensity rainfall events had a median event-average rainfall intensity of 1.1 mm/h (type 1a) and 1.2 mm/h (type 1b), while the moderate-intensity events were characterized by more than twice this median event-average rainfall intensity (3.3 and 2.9 mm/h for type 2a and 2b, respectively). These differences in event-average rainfall intensity between the low- and moderate-intensity event types were statistically significant. The median average three day antecedent sum of precipitation was one order of magnitude smaller for the rainfall events with dry antecedent conditions (2.1 mm and 1.7 mm, for type 1a and type 2a events respectively) than for the events with wet antecedent conditions (27.9 mm and 22.4 mm for type 1b and type 2b, respectively). This difference was also statistically significant. The four rainfall event types also differed distinctly from each other in other characteristic, e.g., the median maximum rainfall intensity during these events and the median duration of the rainfall events (Tab. 1). For all characteristics the IQR was large, reflecting the considerable variability within the four rainfall event types.

#### 2.4. Groundwater Response Time Characteristics

During a typical rainfall event, the groundwater response can be divided into several characteristic phases (Fig. 2). First, there is a delay between the onset of rainfall and the start of the groundwater level rise. In this study, we denote this as the *time to rise*, ( $t_{rise}$ ) and defined it as either the first time step after the beginning of a rainfall event with a positive slope, or the time step with the largest change in groundwater level if the groundwater level was already rising at the start of the rainfall event, which

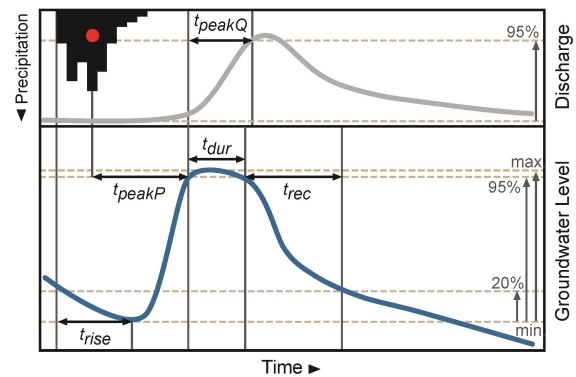


Fig. 2: Schematic groundwater and streamflow hydrograph and timing characteristics as described in Tab. 2. The red dot represents the centroid of rainfall.

was sometimes the case under very wet antecedent conditions. Groundwater responses with an absolute rise smaller than the accuracy of the water level loggers (STS: ca. 0.5 cm and Odyssey ca. 1 cm) were considered as no response. The sum of rainfall that fell between the start of the rainfall event and  $t_{rise}$  is referred to as  $P_{rise}$ .

After the start of the groundwater response, the groundwater level rises to its maximum. For the Alptal catchment this period lasts between less than an hour and up to one or two days, depending on the type of rainfall event. We defined the *time to peak* ( $t_{peakP}$ ) as the time lag between the centroid of each rainfall event (i.e., the time at which 50 % of total rainfall had fallen) and the time that the groundwater level had risen to 95 % of the maximum rise in groundwater level for each event. We used 95 % of the absolute rise (i.e., 95 % of the difference between the groundwater level at the time of first response and the peak groundwater level; see Fig. 2) because it was considered a more robust measure than the peak groundwater level. This was especially the case for sites where the water level first rose quickly and then continued to rise at a much slower rate.

The groundwater table generally remained high for a certain duration. We denoted this time as the *groundwater peak duration* ( $t_{dur}$ ), which was calculated formally as the difference between the time of the 95 % of the absolute groundwater level rise on the rising limb and the corresponding time on the falling limb (called 95 % recession; see Fig. 2).

When water input from the soil surface and upslope areas decreases and drainage exceeds the input at the monitoring site, the groundwater level starts to fall. We defined the *duration of recession*

Tab. 2: Definition of the timing characteristics (see Fig. 2 for a schematic overview of the timing parameters).

Parameter	Definition
$t_{rise}$	Timelag between the start of rainfall and the first response of the groundwater level [min]
$t_{peakP}$	Timelag between the centroid of rainfall and the timing of the 95 % of the maximum groundwater level rise [min]
$t_{peakQ}$	Timelag between the time of the 95 % of the maximum rise in discharge at the catchment outlet and the time of the 95 % of the maximum rise of the groundwater level [min]
$t_{dur}$	Time between the time of the 95 % of the maximum groundwater level rise on the rising limb of the groundwater hydrograph and the corresponding point on the falling limb (called 95 % recession) [min]
$t_{rec}$	Time between the time of the 95 % of the maximum groundwater level rise and the 20 % of the maximum groundwater level rise on the falling limb of the groundwater hydrograph [min]
$s_{rec}$	Mean slope of the groundwater hydrograph between 95 % of recession and 20 % of recession [cm/min]
$P_{rise}$	Sum of rainfall until the start of the groundwater level response [mm]

( $t_{rec}$ ) as the timelag between the time of the 95 % of the absolute rise on the recession limb and the time of the 20 % of absolute rise on the recession limb. The *mean slope of the groundwater recession* ( $s_{rec}$ ) was defined as the difference between the groundwater level at 95 % and 20 % of the absolute rise on the recession limb divided by  $t_{rec}$  [cm/min].

The timing of the groundwater response was also related to the timing of the streamflow response. It is especially informative to investigate the synchronicity between the groundwater and streamflow responses. We therefore define the *timelag between peak groundwater level and streamflow* ( $t_{peakQ}$ ), as the timelag between the 95 % of the maximum rise in discharge at the catchment outlet (catchment C7) and the 95 % maximum rise in groundwater level.

Other groundwater response time characteristics were determined as well but are not reported here, as they were either highly correlated to the selected five time characteristics or were not as robust as the selected characteristics. All response time characteristics and rainfall characteristics were automatically determined for all rainfall events using a script written in R (version 2.14.1; Development Core Team 2005) to guarantee objectivity. The number of events at each site differed because of data gaps (median: 108 events; 25 % quantile: 101 events; 75 % quantile: 121 events; out of a total of 133 events). Because the number of events differed for each well, the relative response frequency for each site was determined as the fraction of events for which a water level response was observed divided by the number of events for which data was available at that site.

## 2.5 Topographic site characteristics

For the analysis of the topographic controls on groundwater response timing, we determined several

topographic indices for each monitoring site based on a re-sampled 6 by 6 meter Digital Terrain Model (DTM) derived from LiDAR data. Other resolutions were tested but 6 meter was considered fine enough to capture the morphologic features within the catchment and coarse enough to avoid the effects of micro-topography, which are of less importance for the groundwater table. *Local controls* are site characteristics of the measurement location, while *upslope controls* denote the properties of the upslope contributing area (Rinderer et al. 2014). The local site characteristics selected for this study were: local slope gradient (Tarboton 1997), local curvature (Evans 1980; Travis et al. 1975) and TWI (Beven & Kirkby 1979). The upslope site characteristics were the size, mean slope, mean curvature and mean TWI of the upslope contributing area. For the delineation of the upslope contributing area, the triangular multiple flow direction algorithm (Seibert & McGlynn 2007) was applied. All indices were calculated using the open source software SAGA-GIS (Conrad 2007). Additional topographic site characteristics were considered in the analyses but are not reported here as they were either highly correlated with the selected indices or not as robust as the selected characteristics.

To quantify the relation between topographic characteristics and the median response time characteristics, the Spearman rank correlation coefficient ( $r_s$ ) was determined. Some plots show the LOWESS regression curve fitted to the median data values as well. The software R (version 2.14.1; Development Core Team 2005) was used to analyze the data. The 0.05 level of statistical significance was used for all analyses. The Mann-Whitney test was used to determine statistically significant differences between the response time characteristics of the four rainfall event types; the Bonferroni method was used to adjust the p-values.



### 3. Results

#### 3.1 Relative response frequency

During most rainfall events, the groundwater levels showed a distinct response to rainfall in a large number of the wells. Half of all sites responded for more than 84 % of all rainfall events but the median relative response frequency was different for the four rainfall event types. The response frequency was 15-20 % lower for the low-intensity rainfall events than for moderate-intensity events (type 1a: 77 %, type 1b: 71 %, type 2a: 93 %, type 2b: 89%; Tab. 3) and this difference was statistically significant. The difference in the response frequency for the events with dry and wet antecedent conditions was small and not statistically significant. The median response frequency was 5-10 % higher for dry antecedent conditions than for moist antecedent conditions, except for sites with a TWI < 4 and low-intensity events. For these sites and events, the median response frequency was more than twice as high under moist than under dry antecedent conditions, which was a statistically significant difference. In general, the response frequency of sites with a TWI < 4 was 10-20 % lower than the response frequency of all other sites, which was a statistically significant difference. The variability in groundwater response frequency was higher for sites with a different TWI than for different rainfall event types (see IQR in Tab. 3)

#### 3.2 Groundwater Response Timing

##### 3.2.1 Groundwater dynamics

The groundwater response of sites with a low TWI (TWI < 4) was delayed compared to the response of sites with a higher TWI (Fig. 3). The difference was

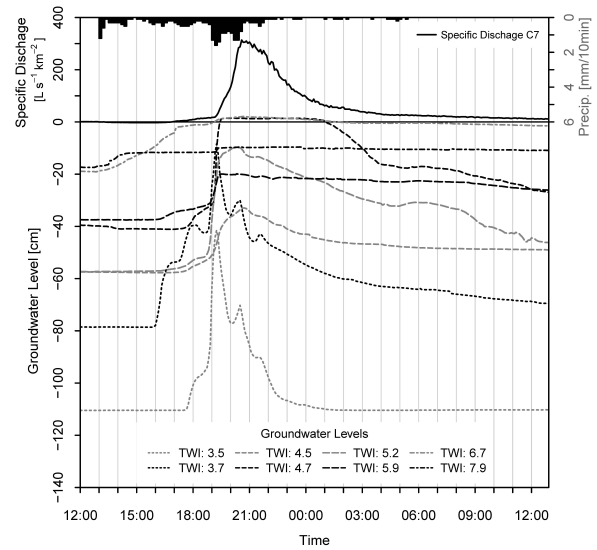


Fig. 3: Groundwater level responses of selected monitoring sites, specific discharge at the catchment outlet (C7) and precipitation for a typical rainfall event (31 May to 1 June 2011). Groundwater monitoring sites with a low TWI are shown with dotted lines, sites with an intermediate TWI with dashed lines, sites with a high TWI with long dashed lines and sites with a very high TWI with dash dotted lines. Specific discharge is shown in the upper panel with a solid line.

on the order of hours and varied for the individual rainfall events. Sites that responded relatively simultaneously, still showed a different response in terms of the gradient, duration and amplitude of the rise (Fig. 3). Groundwater levels rose to the soil surface, not only for sites with a high TWI (e.g., TWI > 6) but also for sites with an intermediate TWI (e.g., TWI: 4-6). While the water level would normally drop soon after the end of a rainfall event for sites with an intermediate TWI, it would generally stay high for several hours to days for sites with a high TWI (Fig. 3).

Tab. 3: Median relative frequency and inter quartile range (IQR) of groundwater responses for different rainfall event types for all sites, for sites with a TWI < 4 and for sites with a TWI ≥ 4. Similar superscript letters indicate when the difference in median response frequencies for sites with TWI < 4 and TWI ≥ 4 are not significantly different). 19 % of the study catchment and 24 % of all wells had a TWI < 4.

Median Response Frequency (IQR)					
Event Type	Intensity	Antecedent Wetness	All Sites	TWI < 4	TWI ≥ 4
All Events	-	-	0.84 (0.70 - 0.90)	0.65 <sup>a</sup> (0.50 - 0.82)	0.85 <sup>b</sup> (0.81 - 0.90)
1a	low	dry	0.77 (0.56 - 0.89)	0.31 <sup>a</sup> (0.22 - 0.56)	0.85 <sup>b</sup> (0.70 - 0.89)
1b	low	wet	0.71 (0.56 - 0.83)	0.65 <sup>a</sup> (0.36 - 0.79)	0.73 <sup>a</sup> (0.61 - 0.83)
2a	moderate	dry	0.93 (0.87 - 0.97)	0.84 <sup>a</sup> (0.57 - 0.94)	0.93 <sup>b</sup> (0.90 - 1.00)
2b	moderate	wet	0.89 (0.74 - 0.95)	0.72 <sup>a</sup> (0.61 - 0.89)	0.92 <sup>b</sup> (0.81 - 0.96)

Tab. 4: Median and full inter quartile range (IQR) of the median groundwater timing characteristics for each well for the different rainfall event types. Similar superscript letters indicate which pairs are not significantly different based on a pairwise Mann-Whitney test and Bonferroni adjusted p-values.

		Rainfall Event Type				
		1a	1b	2a	2b	all
Intensity	-	low	low	moderate	moderate	-
Antecedent Wetness	-	dry	wet	dry	wet	-
Median $t_{rise}$ [min]	median	50 <sup>ab</sup>	78 <sup>a</sup>	20 <sup>b</sup>	30 <sup>b</sup>	35
	IQR	(8 - 162)	(24 - 174)	(0 - 70)	(5 - 78)	(5 - 105)
Median $t_{peakP}$ [min]	median	164 <sup>a</sup>	88 <sup>ab</sup>	80 <sup>b</sup>	58 <sup>b</sup>	75
	IQR	(76 - 273)	(36 - 169)	(25 - 143)	(34 - 101)	(41 - 129)
Median $t_{peakQ}$ [min]	median	-20 <sup>a</sup>	-13 <sup>a</sup>	-25 <sup>a</sup>	-15 <sup>a</sup>	-20
	IQR	(-65 - 95)	(-42 - 59)	(-39 - 28)	(-38 - 25)	(-43 - 18)
Median $t_{dur}$ [min]	median	222 <sup>a</sup>	118 <sup>b</sup>	198 <sup>ab</sup>	135 <sup>b</sup>	145
	IQR	(125 - 293)	(65 - 199)	(80 - 300)	(73 - 190)	(88 - 242)
Median $t_{rec}$ [min]	median	1410 <sup>a</sup>	449 <sup>b</sup>	905 <sup>a</sup>	602 <sup>b</sup>	750
	IQR	(686 - 2078)	(368 - 562)	(483 - 1650)	(360 - 749)	(458 - 920)
Median Response Frequency [%]	median	77 <sup>a</sup>	71 <sup>a</sup>	93 <sup>b</sup>	89 <sup>b</sup>	84
	IQR	(56 - 89)	(56 - 83)	(87 - 97)	(74 - 95)	(70 - 90)
Median $P_{rise}$ [mm]	median	1.2 <sup>a</sup>	1.2 <sup>a</sup>	1.2 <sup>a</sup>	1.3 <sup>a</sup>	1.1
	IQR	(0.4 - 4.0)	(0.6 - 3.1)	(0.5 - 5.1)	(0.5 - 4.4)	(0.5 - 4.1)

### 3.2.2 Time to rise of the groundwater levels

The groundwater level responded to rainfall within minutes to hours. For half of all monitoring sites the median  $t_{rise}$  was less than 35 min, but the variability in median response times among sites was large (IQR: 5 – 105 min). The moderate-intensity rainfall events had the shortest median  $t_{rise}$ . For half of the sites the median  $t_{rise}$  was less than 20 min during the type 2a events and less than 30 min during the type 2b events. During the low-intensity events, half of the sites had a median  $t_{rise}$  less than 50 min (type 1a) and 78 min (type 1b) (Tab. 4). The IQR of the median  $t_{rise}$  of all monitoring sites was more than twice as large for the low-intensity rainfall events than for the moderate-intensity rainfall events (Tab. 4). The difference in the median  $t_{rise}$  between the events with dry and wet antecedent conditions was small and not statistically significant.

The median  $t_{rise}$  for the monitoring sites was correlated to the topographic indices. The  $r_s$  was highest for the mean curvature of the upslope contributing area ( $r_s = 0.82$ ), TWI ( $r_s = -0.81$ ), upslope contributing area ( $r_s = -0.74$ ), mean TWI of the upslope contributing area ( $r_s = -0.66$ ), and local slope ( $r_s = 0.64$ ). The  $r_s$  was lower for the mean slope

of the upslope contributing area ( $r_s = 0.29$ ) and the local curvature ( $r_s = 0.28$ ) (Tab. 5). The median  $t_{rise}$  and the variability in  $t_{rise}$  decreased with TWI for sites with a TWI < 6. For sites with a TWI ≥ 6,  $t_{rise}$  was short and decreased only slightly with increasing TWI, or was constant (Fig. 4). The decrease in median  $t_{rise}$  with TWI for sites with a TWI < 6 was steeper for the low-intensity rainfall events (type 1a and 1b) than for the moderate- intensity events (type 2a and 2b).

The median  $t_{rise}$  was related to the median sum of rainfall until response ( $P_{rise}$ ;  $r_s = 0.98$ ), as well as to the mean and maximum rainfall intensity until response ( $r_s = 0.92$  and  $r_s = 0.96$ , respectively).  $P_{rise}$  was similar and not significantly different for the four rainfall event types (Tab. 4). The amount of rainfall to initiate a response was expected to depend on the soil water deficit and therefore on the antecedent conditions and indirectly on the topography. The median  $P_{rise}$  was indeed correlated to all topographic indices, except for the local curvature (Tab. 5). The median  $P_{rise}$  decreased from > 10 mm to < 1 mm with increasing TWI for sites with a TWI < 6; it was constant or decreased slightly for sites with a TWI ≥ 6 (Fig. 5). Similar results were obtained for the mean and maximum rainfall intensity prior to the start of



Tab. 5: Spearman Rank correlation matrix between the median response characteristics and the topographic site characteristics. For a definition of the response timing characteristics see Tab. 2. Upper right triangle:  $r_s$  values, lower left triangle: p-values. Statistically significant  $r_s$  values are shown in bold font.

	Local slope	Mean slope of the upslope contributing area	Local curvature	Mean curvature of the upslope contributing area	Contributing area	Topographic Wetness Index (TWI)	Mean TWI of the upslope contributing area	$t_{rise}$	$t_{peakP}$	$t_{peakQ}$	$t_{dur}$	$t_{rec}$	$P_{rise}$
Local slope		<b>0.61</b>	0.27	<b>0.61</b>	<b>-0.44</b>	<b>-0.64</b>	<b>-0.43</b>	<b>0.64</b>	-0.05	-0.04	<b>-0.32</b>	<b>-0.39</b>	<b>0.67</b>
Mean slope of the upslope contributing area	<0.001		0.07	0.22	-0.07	-0.24	<b>-0.31</b>	<b>0.29</b>	0.06	0.09	-0.05	-0.13	<b>0.36</b>
Local curvature	0.05	0.64		<b>0.45</b>	<b>-0.48</b>	<b>-0.51</b>	-0.10	<b>0.28</b>	0.02	-0.01	0.06	-0.14	0.27
Mean curvature of the upslope contributing area	<0.001	0.11	<0.01		<b>-0.94</b>	<b>-0.97</b>	<b>-0.80</b>	<b>0.82</b>	0.09	0.09	-0.26	<b>-0.37</b>	<b>0.81</b>
Contributing area	<0.01	0.60	<0.001	<0.001		<b>0.96</b>	<b>0.73</b>	<b>-0.74</b>	-0.13	-0.10	0.23	<b>0.32</b>	<b>-0.72</b>
Topographic Wetness Index (TWI)	<0.001	0.10	<0.001	<0.001	<0.001		<b>0.73</b>	<b>-0.81</b>	-0.09	-0.07	0.27	<b>0.38</b>	<b>-0.81</b>
Mean TWI of the upslope contributing area	<0.01	0.03	0.50	<0.001	<0.001	<0.001		<b>-0.66</b>	-0.19	-0.21	<b>0.29</b>	0.22	<b>-0.65</b>
$t_{rise}$	<0.001	0.04	0.04	<0.001	<0.001	<0.001	<0.001		0.06	0.13	<b>-0.40</b>	<b>-0.35</b>	<b>0.98</b>
$t_{peakP}$	0.73	0.66	0.90	0.54	0.37	0.55	0.18	0.68		<b>0.93</b>	0.23	0.05	0.10
$t_{peakQ}$	0.79	0.51	0.93	0.54	0.50	0.62	0.14	0.35	<0.001		0.11	-0.04	0.16
$t_{dur}$	0.02	0.72	0.68	0.07	0.10	0.05	0.04	<0.01	0.10	0.44		<b>0.50</b>	<b>-0.35</b>
$t_{rec}$	<0.01	0.36	0.34	<0.01	0.02	<0.01	0.12	0.01	0.71	0.77	<0.001		-0.33
$P_{rise}$	<0.001	<0.01	0.06	<0.001	<0.001	<0.001	<0.001	<0.001	0.50	0.27	0.01	0.02	

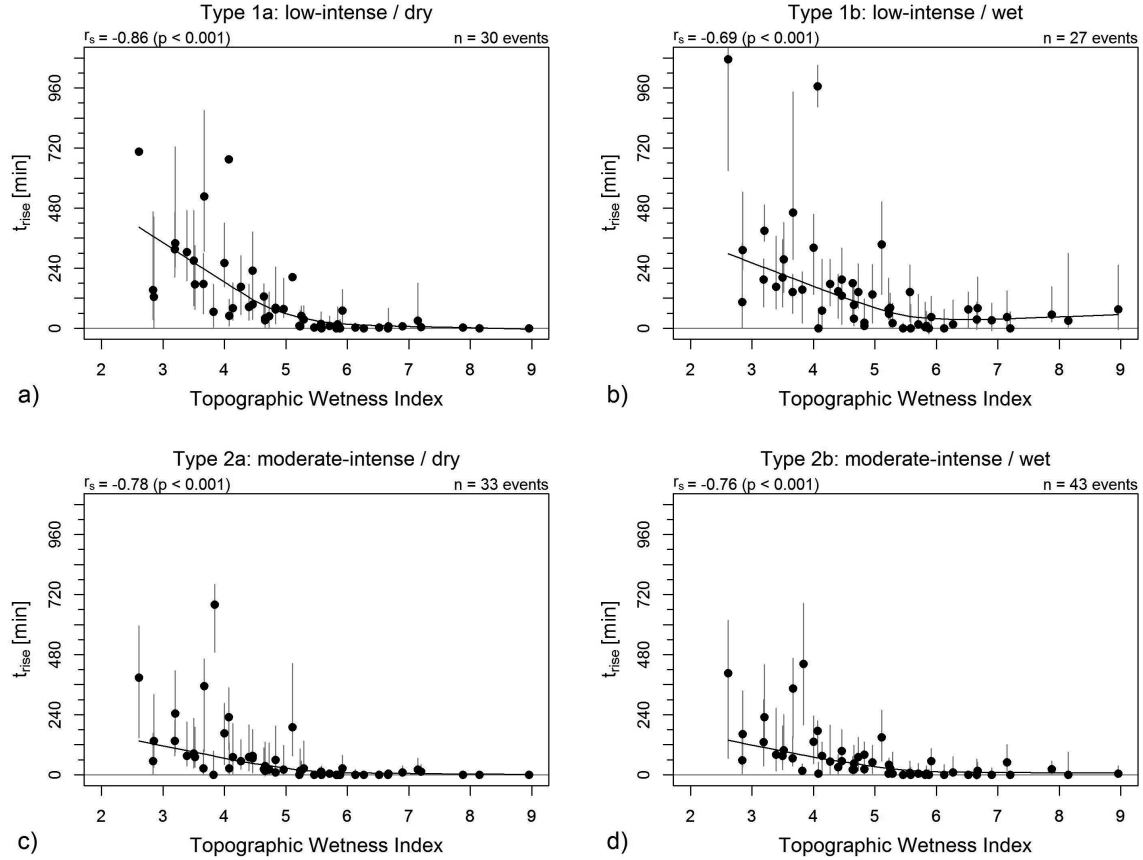


Fig. 4: Time to rise ( $t_{rise}$ ) as a function of Topographic Wetness Index for the four rainfall event types. Grey bar: inter quartile range, dot: median for each site, black line: LOWESS curves fitted to the median values,  $r_s$ : Spearman Rank Correlation Coefficient and associated p-value.

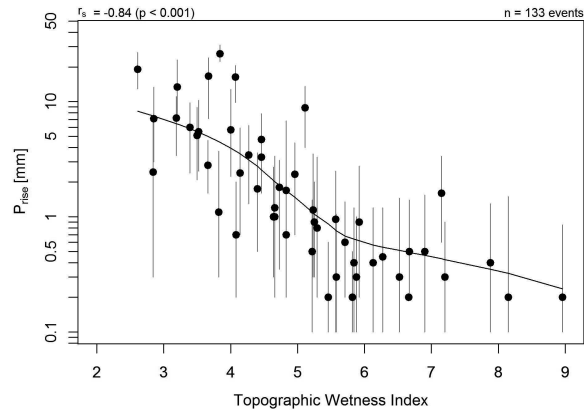


Fig. 5: Sum of rainfall until the start of the groundwater level response ( $P_{rise}$ ) as a function of Topographic Wetness Index for all 133 rainfall events and all sites. Grey bar: inter quartile range, dot: median for each site, black line: LOWESS curves fitted to the median values,  $r_s$ : Spearman Rank Correlation Coefficient and associated p-value.

the rise in groundwater level. However, note that these median  $P_{rise}$  values only include the events for which there was a water table response. This means that for the 1a-type rainfall events fewer events are included in the calculation of  $P_{rise}$  for sites with a  $TWI < 4$  than for sites with a  $TWI \geq 4$ . For the other event types there was no difference in response frequency between the sites with different TWI values.

### 3.2.3 Time to peak groundwater level

In general, groundwater peaks lagged the rainfall centroid. For only 3 of the sites, the median  $t_{peakP}$  was negative (i.e., the 95 % rise occurred before the centroid of the rainfall; see Fig. 6). Variable rainfall intensities during long rainfall events may cause groundwater peaks to precede the rainfall centroid. Differences in  $t_{peakP}$  between the four rainfall event types were small, except for type 1a: the median  $t_{peakP}$  under low-intensity/dry antecedent conditions was less than 168 min for half of the sites, while it was

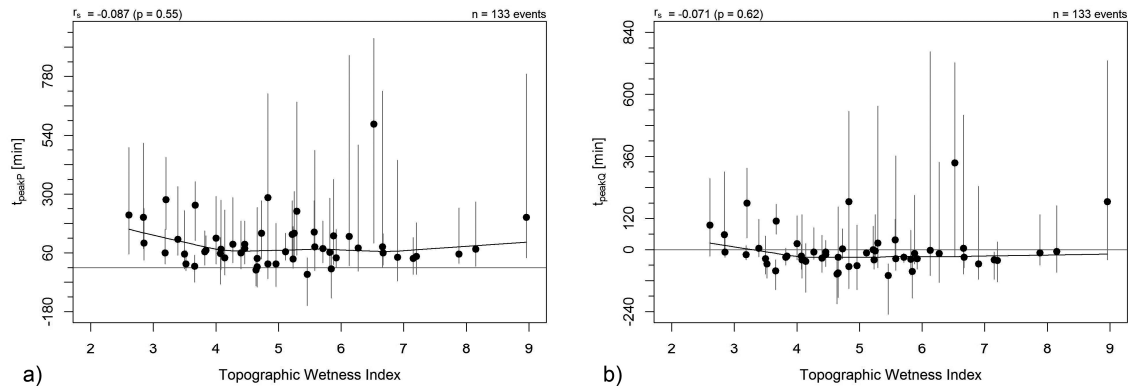


Fig. 6: Timing of groundwater peaks for all 133 rainfall events: a) Time lag between the centroid of rainfall and the time of the 95 % of the maximum rise in groundwater level ( $t_{peakP}$ ) and b) timelag between the 95 % of the maximum increase in discharge and groundwater ( $t_{peakQ}$ ) for all events and all sites plotted as a function of Topographic Wetness Index. The distinct outlier with a  $t_{peakP} > 500$  min and  $t_{peakQ} > 300$  min, is situated in a hollow with an upslope contributing area of  $> 0.1$  ha. Grey bar: inter quartile range, dot: median for each site, black line: LOWESS curves fitted to the median values,  $r_s$ : Spearman Rank Correlation Coefficient and associated p-value.

considerably shorter for the three other rainfall event types (half of the sites had a median  $t_{peakP}$  less than 88 min for type 1b events, less than 80 min for type 2a events and less than 68 min for type 2b events).

The peak groundwater level is expected to precede peak discharge at the catchment outlet (i.e. negative  $t_{peakQ}$ ) if groundwater is the main source of runoff. In general, this was the case, as the median  $t_{peakQ}$  was less than -20 min for half of the sites. However for 13 of the 51 wells, the median  $t_{peakQ}$  was positive (i.e. the 95 % rise in groundwater level occurred after the 95 % rise in discharge). Six out of these 13 sites had a  $TWI \leq 4$  and three out of the 13 sites a  $TWI > 6$ . Differences in the median  $t_{peakQ}$  for the four rainfall event types were small and not statistically significant (type 1a: -20 min, type 1b: -13, type 2a: -25 min, type 2b: -15 min).

When considering all rainfall events, the data supported the assumption that the timing of peak groundwater level would be more influenced by factors like rainfall event characteristics than by topography as the median lagtime to the groundwater peak (median  $t_{peakP}$  and median  $t_{peakQ}$ ) was not correlated to any of the topographic indices (Tab. 5 and Fig. 6). However, for the type 1a rainfall events the  $t_{peakP}$  and the  $t_{peakQ}$  were correlated to the upslope contributing area ( $r_s = -0.43$  and  $-0.44$ ),  $TWI$  ( $r_s = -0.41$  and  $-0.42$ ), the mean curvature of the upslope contributing area ( $r_s = 0.38$  and  $0.41$ ) and the mean  $TWI$  of the upslope contributing area ( $r_s = -0.29$  and  $-0.31$ ). For the type 1b rainfall events, the  $t_{peakP}$  and  $t_{peakQ}$  were correlated only to the mean slope of the upslope contributing area ( $r_s = 0.30$  and  $0.33$ ).

The duration of the peak groundwater level was expected to be a function of subsurface inputs from upslope and thus topography. When all rainfall events were considered together, the median  $t_{dur}$  of a monitoring site was only correlated to local slope ( $r_s = -0.32$ ) and the mean  $TWI$  of the upslope contributing area ( $r_s = 0.29$ ) (Tab. 5). The median  $t_{dur}$  was relatively constant at ca. 120 min for sites with a  $TWI < 4$ , increased up to 180 min with increasing  $TWI$  for sites with a  $TWI$  between 4 and 6 and remained relatively constant at 180 min for sites with a  $TWI \geq 6$ , but this correlation was not statistically significant. Median  $t_{dur}$  was only correlated with local slope,  $TWI$  and mean curvature of the upslope contributing area for the moderate-intensity rainfall events (local slope: type 2a:  $r_s = -0.34$  and type 2b:  $r_s = -0.35$ ;  $TWI$ : type 2a:  $r_s = 0.33$  and type 2b:  $r_s = 0.29$ , mean curvature of the upslope contributing area: type 2a:  $r_s = -0.31$  and type 2b:  $r_s = -0.29$ ).

### 3.2.3 Duration of the groundwater levels recession

The groundwater recession was expected to be slower for sites that receive more persistent water input from their upslope contributing area or that are poorly drained. The median  $t_{rec}$  was indeed correlated to local slope ( $r_s = -0.39$ ),  $TWI$  ( $r_s = 0.38$ ), mean curvature of the upslope contributing area ( $r_s = -0.37$ ) and the size of the upslope contributing area ( $r_s = 0.32$ ) but not to any of the other indices (Tab. 5). The median  $t_{rec}$  increased from ca. 6 hours to 14 hours with increasing  $TWI$  for sites with a  $TWI < 6$ , but the variability was high (Fig. 7). The median  $t_{rec}$  was relatively constant for sites with  $TWI \geq 6$  (14 hours).

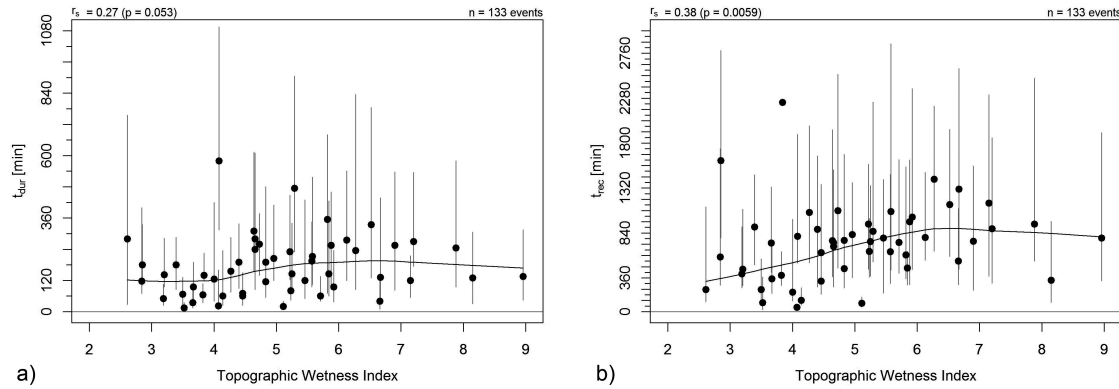


Fig. 7 The groundwater peak duration ( $t_{dur}$ ) (a) and duration of the groundwater recession ( $t_{rec}$ ) (b) for all 133 rainfall events and all sites plotted as a function of the Topographic Wetness Index. Grey bar: inter quartile range, dot: median for each site, black line: LOWESS curves fitted to the median values,  $r_s$ : Spearman Rank Correlation Coefficient and associated p-value.

The median  $t_{rec}$  was longer for the low intensity rainfall events (types 1a, 1b) than the moderate intensity events but the median  $t_{rec}$  and  $s_{rec}$  were not different for the different rainfall event types, except for the event type 1a.

## 4. Discussion

### 4.1 Influence of topography on groundwater response timing

The results show that in the study catchment the timing of the onset of the groundwater rise and the recession are strongly related to topography. The more the flow pathways in the upslope contributing area are convergent (as described by the mean curvature of the upslope contributing area) and the larger the subsurface water inputs from upslope (as described by the upslope contributing area), the faster the groundwater levels respond and the slower they decline. Similarly, the smaller the hydraulic gradient (described by the local slope) and thus the larger the soil wetness (as described by the TWI), the faster the groundwater levels respond and the slower the recession.

Previous studies have not explicitly analyzed the topographic controls on the time to groundwater rise or the duration of the recession. However previous findings appear to agree with our results as groundwater wells near the stream or in footslope locations were well correlated with streamflow or even preceded it (Seibert et al. 2003; Haught & van Meerveld 2011). We can assume that these near-

stream locations had a large upslope contributing area, low slope gradient and high TWI. Upslope wells were not correlated to streamflow and their response lagged behind. Although on hillslopes with more permeable soils, soil depth and bedrock depressions seemed to be more important for the timing of the onset of the groundwater response than surface topography (Penna et al. 2014).

The importance of storage capacity for understanding the time to rise is also corroborated by the strong correlation between the  $t_{rise}$  and the sum of rainfall until the groundwater rise ( $P_{rise}$ ). The strong correlation between  $P_{rise}$  and the topographic characteristics (Tab. 5 and Fig. 5) allows the point measurements to be extrapolated to the catchment scale and thus the parts of a catchment that are likely to respond as a function of cumulative event precipitation to be identified. From this it is, for example, possible to determine, when individual parts of the catchment become hydrologically connected to the stream. According to this functional relation (Fig. 5), wet sites close to the stream and in isolated depressions on hillslopes start to respond on average after 1 mm of cumulative rainfall (44 % of total catchment area). Large parts of the backslopes start to respond after only 1-3 mm of cumulative rainfall (26 % of total catchment area), while after 5 mm of cumulative rainfall 87 % of the catchment exhibits a groundwater response. After 10 mm of cumulative rainfall, the remaining 13 % of the catchment, namely ridges and shoulder locations, are expected to have responded as well (gray shading in Fig. 8). However the large IQR for each site in Fig. 5

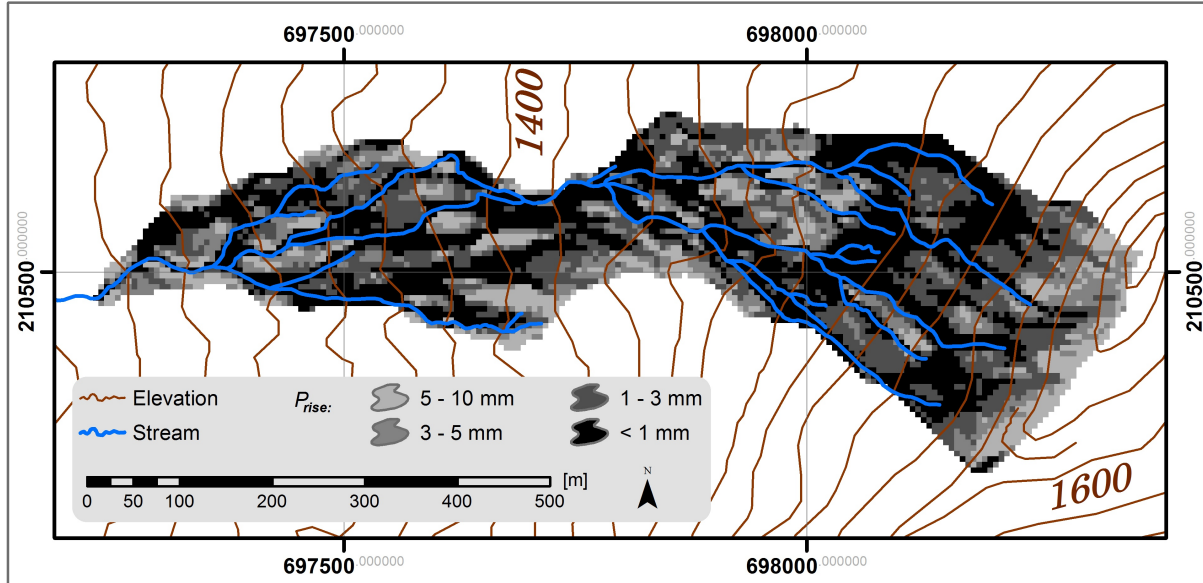


Fig. 8: Spatial distribution of an expected groundwater response after 1 mm, 3 mm, 5 mm and 10 mm of cumulative rainfall based on the relationship median  $P_{rise}$  and TWI in Fig. 5. (background-topographic map: Swisstopo, 123456789)

suggests that this pattern is different for individual rainfall events. The data shown in Fig. 5 also do not account for events that did not cause a response at all. One could expect that this is more likely for sites with a low TWI but response frequencies were generally high and did not show a systematic dependency on TWI, except for type 1a rainfall events. To limit the potential influence of differences in response frequencies we used the median cumulative rainfall threshold determined based on all 133 events to delineate the expected zones of groundwater response in Fig. 8.

Our data did not show a significant correlation of  $t_{peakP}$  or  $t_{peakQ}$  with surface topography when considering all events together. Groundwater peaks normally preceded peak discharge at the catchment outlet but the median  $t_{peakQ}$  was shorter than -20 minutes. In 65 % of all rainfall events, the catchment median groundwater peak occurred earlier than the peak discharge at the catchment outlet. However the large IQR in Fig. 6 also shows that groundwater peaks frequently lagged streamflow peaks by several hours. This is in agreement with other studies elsewhere that have shown, based on end member mixing analysis of hydrochemicals and stable water isotopes, that hillslopes mainly contribute during the recession of streamflow (McGlynn & McDonnell 2003; Burns et al. 2001).

Previous studies have not investigated the correlation between peak lag times and topographic indices explicitly, but have reported that peak-to-peak

lag times vary with soil depth and distance from the stream and therefore with topographic position (uphill-, downhill locations) (Seibert et al. 2003; Haught & van Meerveld 2011; Rodhe & Seibert 2011; Penna et al. 2014). Assuming a different upslope contributing area and TWI for uphill and downhill locations, these findings would not agree with our results. For a more direct comparison we also analyzed soil depth (i.e., well depth, since the wells were installed down to depth of refusal) and distance from the stream (calculated using SAGA GIS) but they were also not correlated with  $t_{peakP}$  and  $t_{peakQ}$ ; nor were they for the four event types separately or for all events combined. Bachmair et al. (2012) reported high spatial variability of peak-to-peak lagtimes, especially during the wet seasons. This agrees better with what we observed in the study catchment, which is wet throughout the year.

The shape of the groundwater peak (as quantified by  $t_{dur}$ ) was more dominated by local drainage than by the subsurface contribution from the upslope. Possible explanations can be that the upslope contributing area was only partly hydrologically connected during events, subsurface flow volumes varied spatially and were not related to surface topography or that drainage affected  $t_{dur}$  more than the variability of the subsurface input.

We expect our findings to be transferable to other humid temperate mountain catchments with low permeability soils and shallow groundwater tables as the topographic characteristics tested in this study

describe physical properties that seem to dominate groundwater flow in these catchments. However, the response time characteristics that we analyzed also showed considerable variability (see IQR shown as grey bars in Fig. 4 to Fig. 7), partly originating from the natural heterogeneity in soil properties (particularly hydraulic conductivity and soil depth) and possibly also from the spatio-temporal differences in rainfall and antecedent wetness within the catchment. It was only possible to identify correlations between groundwater response timing characteristics and the topographic site characteristics by including data for many events at many sites, stressing the importance of large datasets for understanding topographic controls on shallow ground water level dynamics. The large variability also suggests that the timing of groundwater response is controlled by an interplay of static (topography, soil properties, vegetation) and dynamic (rainfall event characteristics and antecedent wetness conditions) controls rather than one dominant factors.

#### *4.2 Influence of rainfall characteristics and antecedent conditions on groundwater response timing*

The time to rise ( $t_{rise}$ ) was influenced more by topographic position than by rainfall characteristics. For dry and intermediate sites ( $TWI < 6$ ), the available storage needed to be filled before the groundwater level would increase, while wet sites ( $TWI \geq 6$ ) seemed to have persistently low storage deficits and therefore responded quickly, regardless of rainfall intensity and antecedent wetness. This suggests that dry and intermediate sites, located mainly on ridges and backslopes, are the zones of highest soil water storage dynamics in the catchment and are most sensitive to differences in rainfall event characteristics and antecedent wetness conditions. This is in agreement with Rinderer et al (2014), who showed for the same study catchment that backslopes with a TWI between 4-6, a local slope between 30-50 % and an upslope contributing area between 200-600 m<sup>2</sup> were the zones of highest variability in median groundwater level. At the same time, our analysis showed that the water level response on backslopes was delayed, while wet sites in footslope locations responded quickly and most likely dominated the rapid streamflow response on the rising limb of the hydrograph. This delay in hydrologic connectivity of the most dynamic groundwater zones in the catchment could be a plausible explanation for the non-linear streamflow response, that is typical for the streams in the Alptal region (Hegg et al. 2006).

The timing of the groundwater peak ( $t_{peakP}$  and  $t_{peakQ}$ ) was dependent on the dynamics of rainfall and did not depend on topographic position. The lagtimes to rainfall were short and similar for the rainfall event types (except type 1a), suggesting that the rainfall input signal propagated quickly to the groundwater regardless of rainfall intensity and event duration. Type 1a events differed from this general behavior since site specific differences in storage deficit, which were related to topography, affected the peak timing of the groundwater levels. The duration of the groundwater peak was also influenced by rainfall intensity, suggesting that only during events with high intensity and potentially also large event sum of precipitation (see Tab. 1), was the upslope contributing area more likely to be connected to the site to enable a persistent subsurface contribution.

The groundwater recession was longer and more variable for dry than wet antecedent conditions. This could partly be an artifact caused by differences in the groundwater amplitude between dry and wet conditions, as the antecedent groundwater levels were lower during dry conditions and the rise in water level during events was larger. Drainage from deeper soil horizons may also be slower due to the lower hydraulic conductivity deeper in the soil profile. In fact, the slope of the recession ( $s_{rec}$ ) was not different for the four rainfall event types.

Previous studies that correlated the average groundwater response across sites on a hillslope with rainfall characteristics have seldom reported actual lagtimes and more frequently reported the number of wells that were activated during events. Nevertheless, our findings are in agreement with these previous studies, as they generally reported the percentage of well activation to be highly correlated with total event precipitation, intermediately correlated with rainfall intensity and weakly or not correlated with antecedent wetness conditions (Bachmair et al. 2012; Penna et al. 2014).

## **5. Conclusions**

The objective of this study was to assess the effect of topography, rainfall and antecedent wetness conditions on groundwater response timing in a 20 ha sub-alpine catchment with low permeability soils. Results of a rank correlation analysis based on data from 51 groundwater monitoring sites for 133 rainfall events suggest that topography is a good predictor for the time to groundwater rise and the duration of the recession but not for the timing of the groundwater peak. Topography controls the time to groundwater

rise by influencing soil drainage, subsurface inputs from upslope, convergence of shallow flow pathways and associated difference in soil water deficits. Topography also controls the groundwater recession by affecting the balance between local drainage and subsurface input from upslope areas.

A rainfall threshold for groundwater initiation existed, which was strongly dependent on topography. The relationships between topographic characteristics and the cumulative rainfall ( $P_{rise}$ ) and time to rise ( $t_{rise}$ ) could allow prediction of the spatial patterns of expected groundwater response zones. This would further enable extrapolation of point measurements to the catchment scale and assessment of the changes of runoff source areas and hydrological connectivity during rainfall events.

Event rainfall and rainfall intensity influenced the time to rise by determining the time needed to satisfy soil moisture deficits and rainfall centroids were shown to control the timing of groundwater peaks. In contrast, the antecedent wetness conditions turned out to play a minor role for the groundwater response timing in this study catchment as groundwater levels are generally high.

We identified topographic indices as good predictors of groundwater response timing, while previous studies in catchments with more permeable soils suggested soil properties and bedrock topography to be more important. From this we conclude that surface topography might play a more important role in determining the variability in groundwater response timing in catchments with low permeability soils and predominantly shallow groundwater tables than in catchments with more transmissive soils (Tromp-van Meerveld & McDonnell 2006; Bachmair et al. 2012; Penna et al. 2014). This would agree with results of Hutchinson and Moore (2000) that hydraulic gradients reflect the surface topography more during periods of high water levels and flow than during periods of low water levels.

Our large dataset allowed us to reveal strong correlations especially between groundwater response timing and topographic predictors, rainfall characteristics and antecedent wetness that may not have been clear from a smaller dataset. The results of this study are expected to be transferable to other catchments with similar topography and soil conditions and allow prediction of groundwater level response in catchments without a dense monitoring network.

## 6. Acknowledgements

We thank Russell Smith and Dan Moore for valuable thoughts on the measurement concept. Special thanks to Benjamin Fischer for sharing data, support and encouragement in the field and numerous fruitful discussions. We thank Tracy Ewen for proofreading the manuscript, Ellen Cerwinka, Jana Dusik, Peter Herrmann, Seraina Kauer, Nadja Lavanga, Claudia Müller, Stephan Müller, Stefan Plötner, Sandra Pool, Paribesh Pradhan, Sandra Schärer, Karl Steiner, Ivan Woodrich, Mirjam Zehnder for field- and lab assistance, Nans Addor for R-tips, Volker Wichmann and Ronald Schmidt for SAGA and GIS-data support, and the Swiss Federal Institute for Forest, Snow and Landscape Research WSL as well as Oberallmeindkorporation OAK-Schwyz, Gemeinde Alpthal and the Canton of Schwyz - Amt für Natur, Jagd und Fischerei, for cooperation.

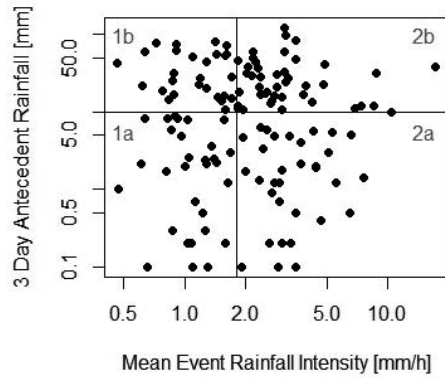
## 7. References

- Anderson, A.E., Weiler, M., Alila, Y. & Hudson, R.O., 2010. Piezometric response in zones of a watershed with lateral preferential flow as a first-order control on subsurface flow. *Hydrological Processes*, 24(16), pp.2237–2247.
- Anderson, M.G. & Burt, T.P., 1978. The role of topography in controlling throughflow generation. *Earth Surface Processes and Landforms*, 3, pp.331–334.
- Bachmair, S. & Weiler, M., 2012. Hillslope characteristics as controls of subsurface flow variability. *Hydrology and Earth System Sciences*, 16(10), pp.3699–3715.
- Bachmair, S., Weiler, M. & Troch, P. a., 2012. Intercomparing hillslope hydrological dynamics: Spatio-temporal variability and vegetation cover effects. *Water Resources Research*, 48(5), p.W05537.
- Barling, R.D., Corporation, W., Moore, I.D. & Grayson, B., 1994. A quasi dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content. *Water Resources*, 30(4), pp.1029–1044.
- Beven, K.J. & Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1), pp.43–69.
- Burns, D. a., McDonnell, J.J., Hooper, R.P., Peters, N.E., Freer, J.E., Kendall, C. & Beven, K., 2001. Quantifying contributions to storm runoff through end-member mixing analysis and hydrologic measurements at the Panola Mountain Research Watershed (Georgia, USA). *Hydrological Processes*, 15(10), pp.1903–1924.
- Burt, T.P. & Butcher, D.P., 1985. Topographic controls of soil moisture distributions. *Journal of Soil Science*, 36(1978), pp.469–486.
- Conrad, O., 2007. *Entwurf, Funktionsumfang, und Anwendung eines Systems für Automatisierte Geowissenschaftliche Analysen*. University of Göttingen.
- Detty, J.M. & McGuire, K.J., 2010. Topographic controls on shallow groundwater dynamics: implications of hydrologic connectivity between hillslopes and riparian zones in a till

- mantled catchment. *Hydrological Processes*, 24(16), pp.2222–2236.
- Development Core Team, R., 2005. *R: A language and environment for statistical computing*, Vienna.
- Dhakal, A.S. & Sullivan, K., 2014. Shallow groundwater response to rainfall on a forested headwater catchment in northern coastal California: implications of topography, rainfall, and throughfall intensities on peak pressure head generation. *Hydrological Processes*, 28, pp.446–463.
- Evans, S., 1980. An integrated system of terrain analysis and slope mapping. *Zeitschrift für Geomorphologie*, Supl. Bd.(36), pp.274–294.
- Fannin, R.J., Jaakkola, J., Wilkinson, J.M.T. & Hetherington, E.D., 2000. Hydrologic response of soils to precipitation at Carnation Creek, British Columbia, Canada. *Water Resources Research*, 36(6), pp.1481–1494.
- Feyen, H., Wunderli, H., Wyder, H. & Papritz, A., 1999. A tracer experiment to study flow paths of water in a forest soil. *Journal of Hydrology*, 225(3–4), pp.155–167.
- Hagedorn, F., Schleppi, P., Bucher, J. & Flühler, H., 2001. Retention and leaching of elevated N deposition in a forest ecosystem with Gleysols. *Water, Air, and Soil Pollution*, 129, pp.119–142.
- Hagedorn, F., Schleppi, P., Waldner, P. & Flühler, H., 2000. Export of dissolved organic carbon and nitrogen from Gleysol dominated catchments—the significance of water flow paths. *Biogeochemistry*, 50, pp.137–161.
- Haight, D.R.W. & van Meerveld, H.J., 2011. Spatial variation in transient water table responses: differences between an upper and lower hillslope zone. *Hydrological Processes*, 25(25), pp.3866–3877.
- Hegg, C., McArdell, B.W. & Badoux, A., 2006. One hundred years of mountain hydrology in Switzerland by the WSL. *Hydrological Processes*, 20(2), pp.371–376.
- Hutchinson, D.G. & Moore, R.D., 2000. Throughflow variability on a forested hillslope underlain by compacted glacial till. , (May 1999), pp.640–655.
- Lana-Renault, N., Regüés, D., Serrano, P. & Latron, J., 2013. Spatial and temporal variability of groundwater dynamics in a sub-Mediterranean mountain catchment. *Hydrological Processes*, p.n/a–n/a.
- McDonnell, J.J., 1990. A Rationale for Old Water Discharge through Macropores in a Steep, Humid Catchment. *Water Resources Research*, 26(11), pp.2821–2832.
- McDonnell, J.J., Sivapalan, M., Vache, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M.L., Selker, J. & Weiler, M., 2007. Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, 43(7), pp.1–6.
- McGlynn, B.L. & McDonnell, J.J., 2003. Quantifying the relative contributions of riparian and hillslope zones to catchment runoff. *Water Resources Research*, 39(11), p.-.
- Mohn, J., Schürmann, A. & Hagedorn, F., 2000. Increased rates of denitrification in nitrogen-treated forest soils. *Forest Ecology and Management*, 137, pp.113–119.
- Moore, R.D. & Thompson, J.C., 1996. Are Water Table Variations in a Shallow Forest Soil Consistent with the TOPMODEL Concept? *Water Resources Research*, 32(3), pp.663–669.
- Penna, D., Mantese, N., Hopp, L., Borga, M. & Dalla Fontana, G., 2014. Spatio-temporal variability of piezometric response on two steep alpine hillslopes. *Hydrological Processes*.
- Rinderer, M., van Meerveld, I. & Seibert, J., 2014. Topographic controls on shallow groundwater levels in a steep, prealpine catchment: When are the TWI assumptions valid? *Water Resources Research*, 50(7), pp.6067–6080.
- Rodhe, A. & Seibert, J., 2011. Groundwater dynamics in a till hillslope: flow directions, gradients and delay. *Hydrological Processes*, 25(12), pp.1899–1909.
- Schleppi, P., Muller, N., Feyen, H., Papritz, A., Bucher, J.B. & Flühler, H., 1998. Nitrogen budgets of two small experimental forested catchments at Alptal, Switzerland. *Forest Ecology and Management*, 101, pp.177–185.
- Seibert, J., Bishop, K. & Nyberg, L., 1997. A test of TOPMODEL's ability to predict spatially distributed groundwater levels. *Hydrological Processes*, 11, pp.1131–1144.
- Seibert, J., Bishop, K., Rodhe, A. & McDonnell, J.J., 2003. Groundwater dynamics along a hillslope: A test of the steady state hypothesis. *Water Resources Research*, 39(1), p.-.
- Seibert, J. & McGlynn, B.L., 2007. A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. *Water Resources Research*, 43(4), p.-.
- Sklash, M. & Farvolden, R., 1979. The role of groundwater in storm runoff. *Journal of Hydrology*, 43, pp.45 – 65.
- Smith, R.S., Moore, R.D., Weiler, M. & Jost, G., 2013. Controls on groundwater response and runoff source area dynamics in a snowmelt-dominated montane catchment. *Hydrology and Earth System Sciences Discussions*, 10(2), pp.2549–2600.
- Tarboton, G.D., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources*, 33(2), pp.309–319.
- Travis, M.R., Iverson, W.D., Gary, H. & Johnson, C.G., 1975. *VIEWIT: computation of seenarea, slope and aspect for land-use planning PSW 11*, Berkeley, California U.S.A.: USDA - Forest Service.
- Troch, P.A., Mancini, M., Paniconi, C. & Wood, E.F., 1993. Evaluation of a Distributed Catchment Scale Water Balance Model. *Water Resources*, 29(6), pp.1805–1817.
- Tromp-van Meerveld, H.J. & McDonnell, J.J., 2006. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, 42(2), pp.1–11.
- Weiler, M., McDonnell, J.J., Tromp-Van Meerveld, I. & Uchida, T., 2005. Subsurface Stormflow. In G. M. Anderson & J. J. McDonnell, eds. *Encyclopedia of Hydrological Sciences*. pp. 1719–1732.



## Supplement Material



Suppl. 1: Division of the rainfall events into four rainfall event types based on the mean event rainfall intensity (threshold = 1.8 mm/h, the mean rainfall intensity that caused a groundwater level response in 10 % of the sites) and 3 day antecedent rainfall (threshold = 10 mm, the mean 3 day sum of precipitation for all events): 1a: low-intensity / dry, 1b: low-intensity / wet, 2a: moderate-intensity / dry, 2b: moderate-intensity / wet.



## STUDY SITES





a)



b)



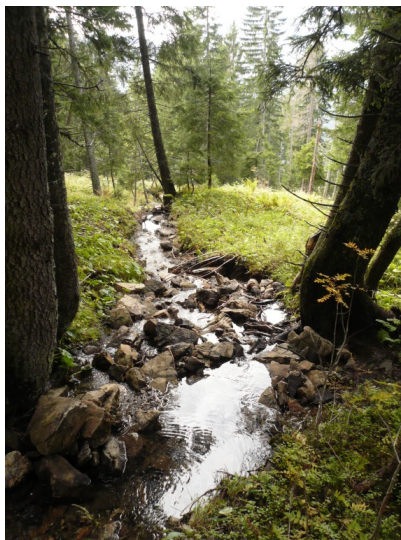
c)



d)



e)



f)



g)

Typical landscape and stream channels in the study catchment: a) wetland, b) saturated area, c) landslide, d) steep hillslopes and light forest, e) stream channel in the upper catchment, f) in the middle part (~12ha) and g) at the catchment outlet(~20ha); (picture: M. Rinderer).





a)



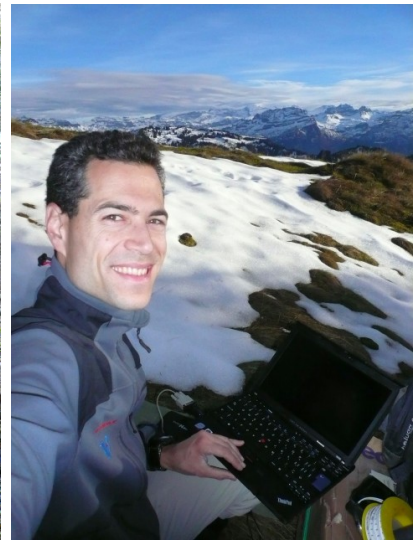
b)



c)



d)



e)



f)



g)

Two of the seven streamflow gauging stations installed in the study catchment (a) v-notch, b) natural crosssection. Site visits and fieldwork with c) Ben, Ilja and Jan, d) Ben with heavy tools, e) myself downloading groundwater level data. Damage of monitoring sites was common, particularly after the winter season: f) broken groundwater well due to soil creep, g) steel stakes of a flume bended by snow creep (picture: M. Rinderer).





a)



b)



c)



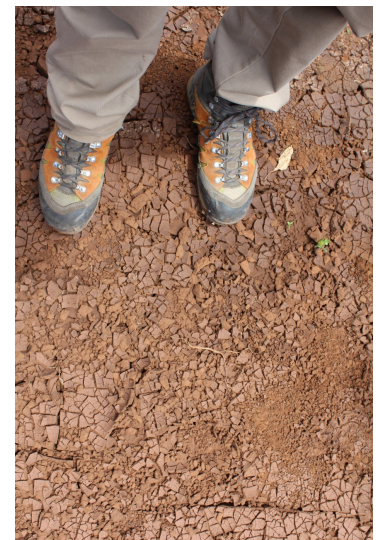
d)



e)



f)



g)

Typical landscape of the study area near Arusha, Tanzania: a) semi-arid lowland with Mt. Kilimanjaro in the background, b) pile of bricks near the village Mungushi, c) flock of cattle in Mungushi with Mt. Meru in the background, d) traditional home near Mungushi, e) “hard” soil moisture measurements using a Time Domain Reflectometry (TDR) device, f) “soft sensing” with fingers and g) with boots; (pictures: T. Siegfried, M. Rinderer).





a)



b)



c)



d)



e)



f)



g)

Qualitative soil moisture assessment test near the villages Mungushi and Kichangani: a) introduction to the qualitative soil moisture assessment scheme b) training participants how to apply the method in the field, c) a PhD-student assessing soil moisture using the qualitative scheme, d) TDR measurements for comparison of qualitative and quantitative soil wetness, e) a local farmer assessing soil moisture using the new scheme, f) local kids enjoying the “soft sensing” test, g) some future local experts on “soft sensing” (pictures: T. Siegfried, D. Müller, M. Rinderer).